University of Alberta

Sensing Film:
A Cognitive Approach to Film Narration and Comprehension

By

Dorit Naaman ©

A thesis submitted to the Faculty of graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

In

Comparative Literature - Film Studies

Department of Comparative Literature, Religion, Film & Media Studies

Edmonton, Alberta

Fall 2000

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59640-0

Canada

University of Alberta

Library Release Form

Name of Author: Dorit Naaman

Title of Thesis: Sensing Film: A Cognitive Approach to Film Narration and Comprehension.

Degree: Doctor of Philosophy

Year this Degree Granted: 2000.

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis not any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

32 Bennett Avenue
Binghamton, NY, 13905, USA

8/18/2000

Sensing Film:

A Cognitive Approach to Film Narration and Comprehension

A Ph.D. Dissertation

Abstract

This dissertation is written within the sub-field of cognitive film theory. It demonstrates some of the benefits of cognitive science research for the understanding of cinematic narration and comprehension. In particular it tries to address a problem in film narratology, namely, that it relies heavily on literary models, and does not adequately describe the particulars of the film medium. This thesis makes a twofold claim: first, film interpretation should be considered not only as a procedure of high order cognitive mechanisms (such as problem solving, memory retrieval, etc.). Instead, interpretation should be considered to be based on the interaction between low order sense and medium specific perception (attention to visual and aural information from the environment), and the aforementioned high order cognitive mechanisms. Second, visual perception and cognition operate differently from language perception and cognition, and these differences affect the construction of the narrative by the spectator. These two claims are demonstrated by a close look at the issues of point of view (as an example of a device in narration), and gap filling (as a form of inference and hypotheses making by a perceiver, and therefore an essential tool of comprehension). The claims of this dissertation are supported by a shot by shot analysis of several scenes from conventional narrative films.

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled *Sensing Film: A Cognitive Approach to Film Narration and Comprehension* by Dorit Naaman in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Comparative Literature - Film Studies.

WR Beard
Bill Beard (Supervisor)

Uri Margolin

Jeff Pelletier

Jerry Varsava

Gary Libben

Edward Branigan
(External Reader)

8/16/2000

Dedication

This thesis is dedicated to my grandfather, Yoram (Konrad) Jacoby, a true intellectual whose life circumstances -- and those of the Jewish people in the twentieth century across Europe and in Israel -- prevented him from exploring his true calling, philosophy.

Acknowledgments

I would like to thank the following people who encouraged me to travel the long path towards the completion of this thesis, and consulted, read, and recommended material that will make this thesis better: Bill Beard, Uri Margolin, Jeff Pelletier, Gary Libben, Jerry Varsava, Edward Branigan, Renée Elio, Marisa Bortolussi, Peter Dixon, Nasrin Rahimieh, Eric Dietrich, Irene Applebaum, Owen Shapiro, Lester Friedman, Jane Wilson, Janet Ould, Peter Bianco.

This thesis would have not been written without the emotional support of the people who shared my life before and during the years of writing this thesis: Yael and Isaac Naaman, Michael Naaman, Heidi von Graevenitz, Andrew Light, Mason Cash, Sally Brackett, Salma Abu-Ayyash, Galit Saporta, Anita Sielecki, David Brown, Carol Fedun, Michael Nahir, Wes Dean, Doris Hambuch, Jennifer Babcock, Jan Olson, Jerry White, Jay Levine, Leslie Heywood, Melissa Zinkin, John Arthur, Amy Shapiro, Jon Mahony, and my many other friends who have always expressed interest in the project, and trusted that I can finish it.

I would also like to acknowledge Arlo, beloved kitty cat, who for the life of him could not understand why I needed a Ph.D., and therefore made sure I took appropriate breaks.

Table of Contents:

Introduction 1 - 2
  A brief personal history of my interest in this project 2 - 6
  Theoretical framework 6 – 10
  Theses 10 - 16
  Thesis outline 16 -19

Chapter 1

What is wrong with film narratology?

A review of the contributions and problematics of applying literary narratology to cinema

  Introduction 20 - 22
  The comparativist project and the case of film and literary narratology 22 - 26
  Film narratology 26 - 29
  Focalization 29 - 33
  Enunciation 33 - 36
  Cinematic enunciation, focalization and point of view 37 - 51
  Summary 51 - 53

Chapter 2

Cinematic implications of cognitive science research on the perception and cognition of images and language.

  Introduction 54 - 55
  Meaningful perception 56 - 62
  Bottom-up and top-down perception 62 - 75
  Language perception 76 - 85
  Visual perception 85 - 88
    The Ventral stream 88 - 93
    Geons 94 - 98

    The Dorsal stream 98 - 109
  Mental imagery and memory 109 - 116
  Summary 116 - 118

Chapter 3

A cognitive approach to film narration

  Introduction 119
  Semiotics 120 - 125
  Semiotic film theory 126 - 139
  Style, or bottom-up perception 139 - 148
  Narration 148 - 171
  A cognitive approach to point of view 171 - 174
  Rambling Rose 175 - 191
  Summary 191 - 192

Chapter 4

The act of perceiving a film: interpretation and gap-filling practices.

  Introduction 193 - 195
  Reading within the text 195 - 203
  Reading in the reader 204 - 212
  What are gaps? 212 - 220
  Cognitive gap filling in theory 220 - 238
  Filling-in filmic gaps 239 - 254
  Summary 254 - 255

Conclusion

  Thesis 256 - 258
  Chapter summary 258 - 267
  Future directions 267 - 271

Bibliography 272 - 283

Appendices

Storyboards (illustrations made by Peter Bianco):

1. *Rambling Rose*(Rose's introduction to Dad) 284 - 292
2. *Rambling Rose* (Initiation to patriarchy) 293 - 309
3. *Dead Poets Society* 310 - 323
4. *Dangerous Liasions* 324 - 332
5. *The Silence of the Lambs* 333 - 351

Illustrations:

1. The Müller-Lyer illusion 72
2. Biederman's decomposition into geons. 95

Introduction

Films overwhelm our senses with images, words, sound effects, and music. It is almost a miracle that when we leave the theatre we can provide a coherent summary of the story. Somehow we manage to process the wealth of sensory information into a linear, causal plot -- a narrative. Moreover, when viewing a conventional narrative film, many of us will summarize the plot in relatively similar terms. But certainly, we do not all "read" every film in the same way, or at least we may disagree about a film's interpretation, overall meaning, and politics.

Recent postmodern theories, as well as reader-response criticism, give an explanation of why we understand artistic texts differently. According to these theories (in grossly generalized terms), the meaning of a text is open; meaning is a product of *how* we read the text, and is influenced by our identity politics and education. But if we accept such theories of meaning, we have fewer resources to explain why we actually interpret conventional cinema generally in the same ways. At the same time, text-oriented theories like formalism, structuralism and semiotics give an account of the production of meaning out of a given segment of a text, but this account is based on textual features and therefore does not allow for differences in interpretation. It is the purpose of this thesis to try and articulate why, despite individual identity differences, we comprehend the narratives of mainstream fictional cinema in much similar ways, that is we agree on plot events, their order, etc. But at the same time, we sometimes interpret the same film in different ways, that is we ascribe certain moral status to the different to characters, categorize the film as belonging to a philosophical or political position, etc. Using research from cognitive psychology I shall give an account of film perception, and then build on this perception to explain the operation of higher cognitive processes in decoding films: those of comprehension, interpretation, and the production of meaning out of textual signs.[1]

Film communicates through the use of images, verbal language, music and sound effects, and each of these stylistic channels of information is cognitively processed in a different way. In this dissertation, though, I will focus only on two of those channels: language and image. I hope that in the future this model could be expanded to music and sound effects as well.

Based on the typology (of perception), I propose a revision of existing models of film comprehension. This approach will account for the ways in which the perception of textual material influences both narration tactics and the spectators' production of hypotheses about the development of the plot. A cognitive understanding of film narration and comprehension will also shed light on the general interpretive practices employed during the viewing of a film. In this thesis I propose a theoretical framework that sheds light on bottom-up perception and top-down cognition, and provides a framework for experimentation on the roles of language and image perception in film narration and comprehension.

A brief personal history of my interest in this project

While working on the shooting script (a breakdown of the script into a detailed shot list) of my student film *High Wire*, I came to draft a pivotal scene. In the film, a married couple pretend to be having an affair with each other in an attempt, among other things, to liven up their marital life. According to the script, the first third of the film keeps the audience thinking of the two as lovers. Then comes the revelation that they are actually married, and from that point on, there is a psychological deterioration of the woman's mental state, to the point that the game has to be stopped. Now, I was coming to plan the shot list of the discovery that the lovers are actually married.

---

[1] The terms: comprehension, interpretation, and meaning production will be defined and explained in detail throughout the dissertation.

In the script, (which I had written a few months earlier), the couple's six year old daughter "catches" the two on the couch (naked under an afghan), after a lovemaking scene. We hear the girl saying "Mommy, I can't sleep," and then we see her at the doorway. I intended for the audience to think the couple was caught by the girl, and then to be surprised as the man gets up to escort the girl back to her room. The scene then cuts to the hallway, where the girl starts a line with the word "Daddy . . ." At this point, I was hoping, my audience would realize that the "lover" is actually the father and husband, and that the couple are playing a game.

Working on the shooting script I had to decide how to let the audience know that the couple were actually married. That is, I had to decide where to shoot from, in how many shots, and of what frame size (long shot, close-up, etc.). Loyal to my personal filmic aesthetics I was interested in shooting the scene in as few shots as possible. Much like André Bazin, I believe in long takes, where the action is developed through the use of depth of space, rather than through editing.[2] In addition to this aesthetic choice, I believed in subtlety and non-redundancy of information: if something was conveyed visually, there was no need to repeat it verbally. I therefore planned to shoot a shot of the couple on the couch from the girl's point of view, then cut to show the girl (from the parents point of view), and then back to the parents, where dad stands up and leaves the frame. We then cut to a point of view shot from the wife's perspective, and we hear the girl's line while seeing both father and daughter from their back, leaving the frame.

---

[2] See André Bazin What Is Cinema, I &II, Berkeley and Los Angeles: University of California Press, 1971.

This plan did not take into account the fact that due to the complexity of the story and the shortness of the film (only 12 minutes in total), the scene had to explicitly mark and emphasize the epistemological shift ("not lovers but a married couple") to the audience. And it is hard to achieve explicitness in a long shot, which is full of details, and is scarce in emotions. Close-ups, on the other hand, would convey much more clearly the emotions and reactions of all involved, thus directing the audience toward a concrete interpretation of the scene. Luckily, my cinematographer (who secretly disagreed with my long shot aesthetics) shot a few close-ups when I was not noticing. In editing, therefore, I had a close-up of the woman's reaction from the couch, but not a close-up of the daughter or the husband. Consequently, the scene is not as clear as it needs to be. While the strongest hypothesis proposed by viewers is the intended one, most audiences tend to feel insecure about the meaning of the scene, some interpreting the whole couch affair as the wife's fantasy, and some, because of the lack of the close-up of the man, even say "her lover looks so much like her husband." Ironically, the portrayal of the psychological state of the heroine benefits from this confusion, as the audience is as lost in this relationship as she is. Despite this small comfort, the experience was frustrating, as an understanding of the twist in the plot is crucial for the understanding of the whole film.

My failure in providing a coherent story in *High Wire* is not atypical for film students. Unlike students, much of the skill of professional directors and editors in narrative cinema is evaluated according to their ability to plan, shoot and edit a film, so as to create a coherent story. Suspense and misleading information are carefully used, and then usually overtly negated (sometimes even repeatedly, or at least through the use of multiple information tracks, i.e., visuals, dialogue, action, etc.), so as to not "lose" the audience to the wrong interpretive track. Indeed, much of the joy of film viewing is in producing the right hypothesis as to what is about to happen, but maintaining a sense of insecurity about that hypothesis until the very end of the film. Should we know exactly what's about to happen we would be bored; should we be totally misled, we would be frustrated. A good filmmaker walks the fine line between the two options, constantly affirming some hypotheses, while canceling others. And much of the production process of a film is devoted to decisions about *when* and *how* to disclose a piece of information, so as to control and direct hypothesis making.

Since the making of *High Wire* I have improved my storytelling skill, and I am now capable of leading my audience where I want the story to go. But since the *High Wire* experience I have been fascinated with the ways in which the structure of a film (or its entire *discours* if one wishes), influences the ways in which we as viewers, derive meaning and pleasure out of a film. When I started to look for theoretical material on the subject, I was disappointed with the level of generality of most writing on the topic. While formalism and semiotics deal with meaning production on the micro level of individual signs, narratology discusses plot and narrative structures mostly in the context of large-scale generalized models. Film narratology, in particular, deals with plot structures, narration, and practices of discourse, but very little attention is paid to the actual source of information, i.e., images, dialogue, music and sound effects. Moreover, the issues of how we process these different channels of information, and how their processing influences our hypothesis making and our understanding of films are very rarely raised. In this dissertation I attempt to use my filmmaking knowledge in order to stay as close as possible to actual textual practices, while attempting to construct a theoretical model of film viewing and understanding. I am particularly interested in the production of meaning out of concrete signs, and its influence on the overall construction of the plot and the film.

Theoretical Framework

Many contemporary academic disciplines are interested in the question of the production of meaning. The range includes fields that focus mainly on the question of meaning itself, like philosophy of language or mind, and linguistics. But at the same time cognitive and computer sciences are interested in the implication of knowledge about the ways humans produce meaning, for computer technology and artificial intelligence. Critical thinking in the humanities has also led to an interest in the artistic devices and functioning of signs and their correlative interpretive processes. The fields of semiotics, structuralism and reader-response criticism are preoccupied with the production of meaning out of artistic texts. It is the purpose of this dissertation to conjugate some of these fields, in an attempt to propose new perspectives on narrative comprehension. But before stating the specific claims of this thesis, I need to describe in greater detail the fields that are relevant to film theory. Film theory is a young field, which emerged mostly out of the well-established disciplines of literary studies, linguistics and to some degree philosophy. In its early days, film (then called the movies, pictures, etc.) was considered to be mostly an entertainment medium, and film researchers tried to qualify it as a serious art. The reliance on well respected literary studies, and the attempt to apply linguistic models to film analysis were done partially in order to redeem film's respect as an art form, and to establish film studies as a creditable field of inquiry. But the application of literary and linguistic models to film is problematic. While literary theorists and linguists study verbal messages, film communicates with images as well as with verbal languages. The linguistic models applied, and the attempt to qualify film as a language (or at least a language system), neglected to treat film in its medium specific conditions (i.e., images, and sounds). Film narratologists and cognitive film theorists, for their part, often struggle to address the multiplicity of sensory information in film, and as a result they produce models that do not account for the complexity of the experience. In the following pages I provide a brief review of the different fields involved in the study of film comprehension and interpretation, and I map them according to their relevance to this study.

Film and literary criticisms have been pre-occupied with the relationship between the perceiver and the text, or how meaning and content are to be understood. The early half of the century saw the emergence of formalism, structuralism and semiotics.[3] Semiotics and formalism regard the text as a system of signs, which can be understood as a relationship between a signifier (a communicative practice) to a signified (an object, or a concept). Structuralism took into account the ways in which the text is organized into meaningful clusters of information. According to these theories, if one is able to decode all the signs (verbal, image based, metaphoric, etc.) and to formulate the structural devices employed, one is able to *understand* the text. In other words, the text is the locus for meaning, which is communicated through sets of signs and formalistic devices. But Reader-response Criticism has since pointed out that readers tend to form a relationship to the text, and that this relationship is not universal for all readers, but is particular to each reader, and is dependent on ideology and on the identity of the reader.[4] This identity is based on factors such as race, gender, age, religion, nationality, etc. Meaning, according to this approach, is negotiated between each and every reader and the text, and cannot be inherent to the text. In addition, post-structuralism, post-modernism and deconstruction theories challenged the notion that the meaning is *in* the text, by shaking the (previously assumed) tight relations between the signifier and the signified.[5] If the signifier does not automatically refer to a concrete signified, there is a gap to be filled by the interpretive process.

---

[3] For literary and film formalism see the works of Shklovsky, Eikhenbaum, Tynianov, Pudovkin and Eisenstein. For Structuralism see the works of Jakobson, Levi-Strauss, Todorov, Barthes and Genette. For cinematic semiotics see the works of Eco, Metz and Pasolini.

[4] See Hall, Fish, Fiske, Bennett, and Glasgow Media Group.

[5] See Derrida, Foucault, Kristeva, Barthes, and, specifically for film, Ropars-Wuilleumier, Ryan, Ulmer, Brunette, Wills, and Heath.

And the interpretive activity is a subjective and relative one, dependent on the specific inclinations of an individual.

Recent film and literary narratology too, urge us not to look at the processing of "bottom-up" (textual and structural information) only. It is important (some argue, even more important) to account for other activities of the brain, the ones generally called "top-down" processes. Top-down operations are those that are associated with beliefs and expectations we bring to the perception of the text (such as generic expectation, general knowledge, etc.). In addition, the text leaves much information to be provided by the imagination of the perceiver. Dramatically crucial information is often strategically not provided, so as to create a dramatic effect, one that will lead to suspense, surprise, and greater emotional involvement of the perceiver. These gaps evoke the perceiver to apply a range of mental activities in order to bridge them with information, or hypotheses.[6] In the case of redundant information, one uses one's everyday knowledge about the world, so as to imagine what had happened during the temporal or spatial gap, in order to achieve a sense of closure. In the case of a generic film, one uses one's knowledge about the genre in order to provide this missing information. But in the case of dramatically crucial information, the perceiver is not only imagining, but is also producing hypotheses about what is missing and how it will affect the plot line. These hypotheses are complex mental activities which involve memory retrieval, problem solving techniques, and the psychological desire to achieve narrative closure. Inference making and the above mentioned high order cognitive activities have been studied quite extensively in the past few years, but the research refrains from determining the relationship between the perception of "bottom-up" information and the "top-down" processes.[7] It is the purpose of this thesis to explore the particular mechanisms of film perception and cognition and how the "bottom-up" and the "top-down" interact to produce narrative comprehension.

---

[6] For a comprehensive discussion of gap filling in literature see the work of Iser and Ingarden.

Cognitive film theorists, for their part, have been focusing their attention on the medium as mostly a visual medium. The photographic moving image seems “real” by a series of technical manipulations (involving the camera and the projector), and the audience is aware of those manipulations. Yet, at the time of perception, perceivers are consumed into the realistic experience of what unfolds on the screen, as if these events (which were shot sometime in the past, and then edited) are actually occurring in front of their eyes. Much of the cognitive film literature is devoted to an exploration of the nature of the filmic illusion.[8] But these theorists often treat film as if it is comprised of images only, and ignore the impact of the dialogue, music, and sound effects. A full cognitive account of film narration needs to address the medium in its full complexity.

---

[7] A partial list includes: David Bordwell, Narration in the Fiction Film, (Madison: University of Wisconsin Press, 1985), and Making Meaning: Inference and Rhetoric in the Interpretation of cinema, (Cambridge, Mass: Harvard University Press, 1989); Edward Branigan, Narrative Comprehension and Film, (Routledge, 1992); Ian Douglas, Film and Meaning, (Continnum Publications, 1988); David Boyd, Film and the Interpretive Process, (Peter Lang, 1989), and Warren Buckland, The Film Spectator, (Amsterdam University Press, 1995).

[8] See Gregory Currie, Image and Mind: Film, Philosophy, and Cognitive Science, New York, NY: Cambridge University Press, 1995; Richard Allen, Projecting Illusion: Film Spectatorship and the Impression of Reality, New York, NY: Cambridge University Press, 1995; Trevor Ponech "Visual Perception and Motion Picture Spectatorship," Cinema Journal 37:1, 1997, pp. 85-100.

Theses

The term "narrative" represents a complex relationship between a text and the plot it tells. While the plot is embedded into the text, the text provides much information (like description of places, characters, etc.) which is not necessarily directly relevant to the construction of the plot. Gregory Currie claims that:

> A text is a narrative in virtue of the story it tells. A text is a sequence of words or sentences, something which can be uttered by someone – an author perhaps – on a particular occasion. A story is a set of propositions to the effect that this happened, and then that happened, that this happening caused that to happen, that character A was involved in such and such a way in initiating causation, that character B was affected thus and so.[9]

The narrative, then, is the product of a mental activity that the perceiver of the text performs in order to comprehend the story events. And the story is a re-organization of textual information into a cause-and-effect chain of propositions. This view of narrative is common, and designates the agreement that the narrative is a product of high order cognitive activities (such as re-organization of textual material into cause-and-effect order), which result in organization into propositional sets. This model is sensible when we discuss verbal narratives, as language lends itself easily to propositional arrangement of logical relations. But in the case of film, this narratorial model presents numerous problems. Jacob Lothe points out that:

---

[9] Gregory Currie, Image and Mind: Film, Philosophy and Cognitive Science, NY: Cambridge University Press, 1995, P. 232.

> . . .although film communication clearly has points of contact with verbal communication, the film medium is very different from the verbal form of communication we meet in narrative texts. [. . .] the narrative communication model refers to language and not to film.[10]

Lothe alludes to a separation between the discussion of narrative as a high order cognitive product, and the discussion of the film experience, which is not just verbal and propositional. Tom Gunning, too, asserted that:

> The concept of narrativization focuses the transformation of showing into telling, film's bending of its excessive realism to narrative purposes.[11]

Gunning here postulates the idea that the visual "showing" of narrative events is somehow translated into "telling" or a verbal narrative. That is, the visual material cannot tell a story in its own right, but has to somehow be translated to a linguistic medium in order to become a narrative. And David Bordwell claims that:

---

[10] Jacob Lothe Narrative in Fiction and Film: An Introduction NY: Oxford University Press, 2000, P.13.

[11] Tom Gunning D.W. Griffith and the Origins of American Narrative Film, Urbana: University of Illinois Press, 1991. P.17

> Viewing is synoptic, tied to the time of the text's presentation, and literal; it does not require translation into verbal terms.
>
> Interpreting (reading) is dissective, free of the text's temporality, and symbolic; it relies upon propositional language.[12]

Bordwell here makes a distinction between film perception and interpretation, and he claims that the construction of the narrative (interpretation) operates separately from the perception of bottom-up filmic information. Janet Staiger, on the other hand, argues with Bordwell when she says:

> I disagree with Bordwell's belief that viewing or perceiving can be separated, except in the most theoretical way, from interpreting or reading. I believe that interpretational propositions inform perception and viewing.[13]

Staiger is a reception studies theorist, and she strongly believes that perception is governed by top-down belief operations, that is, that the very perception of low-level information is always ideologically biased. I will perform a more elaborate critique of both Bordwell's and Staiger's positions in chapter 4, but for the time being I would like to point out that neither one specifies the relationship between perception of bottom-up information and the high order cognitive operations which are both required for the construction of the narrative.

---

[12] David Bordwell, Narration in the Fiction Film, op. cit. P. 30.

The first claim of this thesis is that film interpretation should be considered not just as a procedure of high order cognitive mechanisms (such as problem solving, memory retrieval, etc.). Instead, narrative comprehension and interpretation should be considered to be based on the interaction between bottom-up sense perception and those high order cognitive mechanisms mentioned above. In the third chapter of this thesis I provide a few specific examples of how this interaction between perception and interpretation works, particularly when it comes to narratorial tactics. I will focus on point of view as a (very effective) narratorial device that exemplifies how crucial bottom-up perception is, to narrative comprehension. In particular, I will analyze two scenes, one from *The Silence of the Lambs,* and the other from *Rambling Rose.*

The second claim of this thesis is that visual perception and cognition operate differently from language perception and cognition and that these differences affect *how* and *what* we remember of that information, and consequently also affect such high order operations as the construction of the narrative. Generally speaking, while natural languages are highly coded (comprised of arbitrary signs which refer to signifieds by social convention), images are very specific and particular, and they require none or little decoding, in order for us to understand them. The coded, language based information is compacted for storage in the brain, and cognitive scientists argue about the scale and architecture of

---

[13] Janet Staiger, Interpreting Films: Studies in the Historical reception of Cinema.

these storage units.[14] But images are processed and stored differently than language sequences, using both propositional sets, prototypical scenarios, and holistic depictive representations. The most direct effect of these perceptual and cognitive differences is on the construction of the memory of this information: what is highly codified and categorized is more likely to be consciously remembered than what is stored as is (an image). This difference in memory storage of images and language, in turn, has an effect on the ways we come up with inferences, produce our hypotheses and fill in narrative gaps. It is commonly agreed upon that the narration of a film is not complete without the active participation of the perceiver. The film rarely provides all the information needed to understand the plot. It usually refrains from giving some information that is dramatically crucial for the understanding of the plot, thus encouraging the viewer to produce hypotheses about what happened/is about to happen, and to fill in narrative gaps based on those assumptions. Upon closure, some of these hypotheses are to be confirmed, some negated, and some altered, and usually the gaps would be filled by the text. Our

---

Princeton: Princeton University Press, 1992. P. 64.

[14] Marvin Minsky uses the concept of frames, with slots for new information (see "A Framework for Representing Knowledge" in P.H. Winston (ed.) The Psychology of Computer Vision, NY, McGraw Hill, 1975. PP. 211-277). Schank and Abelson discuss scripts, which are sequences of already known experiences (see Scripts, Plans, Goals and Understanding, Hillside, NJ: Erlbau, 1977. Ch. 3, p. 52-68). Jerry Fodor talks about a much larger scale unit, the module, which covers all of the syntactic or the semantic operations (see "Precis of the Modularity of the Mind" in Behavioral and Brain Sciences, 8:1-5, 1985). McClelland, Smolensky, Rumelhart, Hinton and others refute the hierarchical nature of all the previous architectonic models, and instead they propose the idea of parallel distribution processing (see "Schemata and Sequential thought processes in PDP Models," in Parallel Distributed Processing, vol.1: 3-40, 1986).

interest in narrative fiction film arises partly because of our involvement in the production of the narrative of the film. My claim here is that when encountering a narratorial gap, we use all perceived information (images, dialogue, music) to produce our hypotheses. But we are more consciously aware of the hypotheses that were produced based on highly codified information (such as language), while we may not be able to articulate why we make certain hypotheses which are based on less codified information (such as images). Similarly, because we are less consciously aware of visual information processing, we may assume a gap exists, while in effect, the information is provided by the image channel of the text. I shall call these gaps illusory gaps, given that the inference is made based on actual perceived information, and not on hypotheses production. In the fourth chapter I will give examples of both real and illusory gaps, and I will provide an account of cognitive gap filling. The filmic examples I will analyze are from *Dead Poets Society* and *Dangerous Liasions.*

Thesis Outline

The first chapter discusses the current problems in film narratology, and suggests that a cognitive approach can resolve some of these problems. In particular, I look at the problems that arise from the fact that film narratology relies so heavily on literary narratology, and is therefore somewhat blind to the medium specific attributes of the film experience (i.e. to the fact that film is also told visually, and not only verbally). I examine the concepts of enunciation, focalization and point of view as specific examples of the problematics of medium transference of theoretical terms. Generally put, in literature, enunciation refers to a speaking agent, focalization to an experiencing agent, and point of view to the actual restrictions on what one sees (often a device in focalization). But in film these neat categories blur. Filmic enunciation may not be verbal, and focalization may be conveyed not necessarily through point of view structure. In this chapter I show that filmic enunciation, focalization and point of view operate in multiple ways via linguistic, musical and visual means, and create complex and sometimes contradictory narratives.

In the second chapter I review research from cognitive psychology that helps establish the ground for a cognitive analysis of film narration and comprehension. This chapter is a wide overview of perception, cognition, and memory storage, of both visual and verbal information. I will first discuss the concept of meaningful perception and then talk about the respective differences in the perception of language and visual material. The differences in low-level perception affect cognitive processing at high levels, such as categorization and memory storage. While I give numerous examples from different films, and focus in greater detail on the opening scene of *Once Upon A Time In The West,* the filmic examples should be read just as illustrations of the cognitive findings. The full application of the material from this chapter to the issues at hand is only performed in the third and fourth chapters.

The third chapter discusses narration. The first part of the chapter reviews the semiotic and structuralist models of narration, and points out the problems with such a linguistic based approach. I will then review the work of several cognitive film theorists, particularly as it pertains to narratology, and suggest a cognitive extension of Edward Branigan's model of narration. The extension accounts for the actual channel of information (verbal or visual), and particularly focuses on issues of enunciation, focalization and point of view. Through numerous filmic examples these cognitive concepts are grounded in a detailed analysis.

The fourth chapter discusses the cognitive activity of the perceiver, namely comprehension and interpretation. The chapter starts with a review of reception studies theories, their contributions to the debate on the activity of the perceiver, and the limitations of present theories. I then propose a cognitive account of comprehension and interpretation, one that takes into account both bottom-up perceptions and top-down assumptions and beliefs. In the rest of the chapter I focus on gap-filling practices as one example of textual moments that invite the perceiver to come up with hypotheses about what has happened, or about to happen. These inferences, I claim, are guided by textual information, and are sensitive to the track that this information was delivered in (i.e. visual or verbal).

To sum up, through a discussion of point of view and enunciation (as modes of narration- ch. 3), and gap filling (as a mode of interpretation- ch. 4) I will show in this dissertation that: (1) interpretation of film should be considered to be based on the interaction between bottom-up sense perception and high order cognitive operations; and (2) that visual perception and cognition operate differently from language perception and cognition, and that these differences affect the activity of the construction of the narrative by a perceiver.

Finally, in my conclusion I try to outline some ways in which the ideas I have explored in the thesis might receive further application in film narratology. In addition, I suggest some areas in which such a cognitive approach to cinema can enrich and improve discussions of film theory in general.

Chapter 1

What Is Wrong With Film Narratology?

A review of the contributions and problematics of applying literary narratology to cinema.

Introduction

This chapter describes the current state of the young discipline of film narratology. In order to focus the discussion of this large field, and exemplify some of the field's problems I will examine the concepts of *point of view, enunciation* and *focalization.* In this chapter I claim that when applied to a discussion of cinema, these literary terms are borrowed without a careful consideration of the complex nature of the filmic medium and its communication systems. An application of these literary terms to the discussion of film is therefore inappropriately reductive, and needs to be seriously re-evaluated. In the following chapters I will offer a cognitive model of film narration and comprehension, one that is hopefully able to solve some of the problems outlined below.

Narratives are not merely an artistic product; they are a phenomenon common to many aspects of our lives such as our cultural and personal histories, everyday events we tell each other, news items, and more. Narratives have been studied as sets of functions, as plot structures, and as a form of discourse; additional attention has been focused on reception and on the cognitive aspects of perception. In the arts, narratives have been explored and studied since the early days of drama, and the theoretical frame applied varies with the historical period.

Gerald Prince defines narrative as:

> The recounting (as product and process, object and act, structure and structuration) of one or more real or fictitious events communicated by one, two or several (more or less overt) narrators to one, two, or several (more or less overt) Narratees.[1]

Regardless of the medium then, a narrative incorporates an act of narrating, an event being narrated and an implied audience. The issue of the addressee -- whether a real person or a fictional narratee -- becomes even more crucial when one comes to discuss narration. As Edward Branigan notes, "narration refers not to the story itself but to the *knowing* of the story."[2] The emphasis here is not only on the source of the information (the narrator), who controls "when" and "what" is to be distributed, but also on a perceiver, who makes a meaningful story out of the act of narration. Whether one wants to perform a structural, phenomenological, or reception-based analysis of narratives, this analysis is anchored in the epistemological product attained by the perceiver.

The Russian Formalists drew a distinction between *fabula* and *syuzhet*. The *syuzhet* refers to the story events as they are organized in the text in a linear progression, though not necessarily in causal relations. The *fabula* is a construct, a re-organization of the *syuzhet* into a causal chain of story events in the right temporal order. The *fabula* then is a construct in the mind of the perceiver, a re-telling of the narration into a coherent story. In his discussion of *fabula* and *syuzhet* David Bordwell concludes that "the fabula is thus a pattern which perceivers of narratives create through assumptions and inferences."[3] But it is clear that the *fabula* is never independent of the *syuzhet*. The order and nature of the narration influences the production of assumptions and inferences on the side of the perceiver.

---

[1] Gerald Prince, Dictionary of Narratology, University of Nebraska Press, 1987, p.58.

[2] Edward Branigan, Point of View in the Cineam: A Theory of Narration and Subjectivity in Classical Film, Mouton Press, 1984. p.2.

In the following pages I will demonstrate the relationship between the film text as a set of structural devices (exemplified by focalization, enunciation, and point of view), to the construction of the *fabula* by the perceiver. But first, I need to discuss the comparativist project between literature and film in general. Film narratology is a young sub-discipline, which emerged out of the well-established field of literary narratology. But the transition from literature to film is often done without careful consideration of the specific material aspects of the medium (i.e.,, communication via images and sounds vs. verbal language in literature). After the general account of the comparativist project, I will discuss the literary concepts of focalization and enunciation, to be followed by a discussion of how these concepts were borrowed into film narratology.

The Comparativist Project and the Case of Film and Literary Narratology

In the *Republic*, Plato distinguishes between two conceptions of narration, telling and showing (diegetic or mimetic). Diegetic narration consists of verbal activity, a telling, (which can be done in the voice of the narrator, the poet, or in the voice of one of the characters). Mimetic narration is based on a dramatic presentation, a showing or a spectacle.

---

[3] David Bordwell, Narration in the Fiction Film. Madison: University of Wisconsin Press, 1985. p.49.

The distinction between showing and telling has been much debated throughout history. At the turn of this century Henry James called for writers to write in the mode of showing alone, so as to let the story present itself to the reader, rather than telling or summarizing the events. In the 1960's, Wayne Booth and others revived the glory of "telling" as the preferable mode of literary transmission. Edward Branigan summarizes Booth's categories of telling and showing. Concepts like narration, mediation, summaries, description, ideas, and language are associated with telling. Showing, on the other hand, is associated with imitation, immediateness, scenes, drama, presentation, images, and pictures.[4] Many theorists articulated the differences between showing and telling, and I shall give two examples. Marie-Laure Ryan claims that

> The "mimetic stratum" of the work is not experienced as language but directly as world: the reader does not simply conclude "the narrator says P" from the narrator's mimetic statements, he also derives "P is the case" and regards P as an unmediated fact of the real world of the fictional universe.[5]

This description of the mimetic aspects of literature is obviously very appealing to cinematic theory of narration. Similarly, André Gauderault describes cinematic narration and monstration, a form of telling and showing, and he claims that:

---

[4] Edward Branigan, op. cit., p.191.

> The "unipanctuality" to which the monstrator is bound prevents it irreducibly from modulating the temporal flow of the narrative. . . It is because the monstrator, any monstrator, clings so closely to the immediacy of the "representation" that it is incapable of opening up this gap in the temporal continuum. . . Only the narrator can sweep us along on its flying carpet through time.[6]

But many objections have been proposed to the distinction between showing and telling, and I would like to quote here Gerard Genette, who dismisses the whole debate as irrelevant for verbal narratives.

> . . . in contrast to dramatic representation, no narrative can "show" or "imitate" the story it tells. All it can do is tell it in a manner which is detailed, precise, "alive," and in that way give more or less the *illusion of mimesis* -- which is the only narrative mimesis, for this single and sufficient reason: that narration, oral or written, is a fact of language, and language signifies without imitating.[7]

Genette calls attention to the abstract and arbitrary nature of all natural languages, whereby a social contractual agreement, a set of letters (organized as phonemes and morphemes) are agreed upon to represent, and thus communicate an object. The letters d-o-g do not imitate in any

---

[5] Marie-Laure Ryan, "Fiction as a Logical, Ontological and Illocutionary Issue," Style 1984, 18:2, 124.

[6] André Gaudereault, "Narration and Monstration in Cinema" Journal of Film and Video 39 Spring, 1987, p. 32.

[7] Gerard Genette, Narrative Discourse: An Essay in Method, trans. by Jane E. Lewin. Cornell University Press, 1972. p.164.

visual way the animal that barks, but it signifies that animal for English speakers who long ago implicitly agreed that d-o-g refers to that particular animal. I will return to Genette's objection to the distinction between showing and telling at the very end of this chapter. However, it is clear that the filmic medium incorporates both showing (images, action, drama) and telling (narration and voice-over), and the way the two modes interact deserves attention. I shall return to this distinction when discussing focalization and enunciation in cinema.

Since the early days of film criticism, film theorists, like Eisenstein or Arenheim have referred to issues of narration and narrative structure in their writings. However, film narratology as a coherent discipline has only started to formulate itself in the past fifteen years. As a result, most of the terminology used until recently was borrowed from literary narratology and applied to film, sometimes without a careful consideration of the differences between the two media. Moreover, some of the terminology used by literary narratologists referred to visual aspects of stories -- "showing," "perspective," "point of view" -- which in turn seemed to ease the transition from literary to film narratology. But while literature is a linguistic medium, film incorporates visuals, music, sound effects and natural languages. Using Genette's objection to literary "showing" we can now see that the visually oriented terminology of literary narratology refers to *description* of space, while film actually produces a visual space, by use of a different signifying system, i.e., photography. The multiplicity of channels of communication in film, the variety of signifying systems, and their effects on the perception of the narrative need to be addressed in a comprehensive theory of film narratology. Thus when narratologists like Seymour Chatman try to articulate film and literature as analogous, the result is problematic at best, and yields an insufficient theory of both.[8]

But some concepts from literary narratology are of course very useful in discussing film, and theorists like David Bordwell, Edward Branigan, David Allan Black, and Jeffrey Rush have successfully adapted Genette's, Chatman's, and others' concepts to address film issues. I shall come back to this work later, and in the meantime let me say that as long as the application of concepts from literature to film is careful, with a consideration to the peculiarities of each medium, the analogy is valid and can be useful. In the next sections I shall briefly describe film narratology, and then focus on the application of literary focalization and enunciation to a discussion of film.

Film Narratology

The two main figures in contemporary film narratology are David Bordwell and Edward Branigan. Bordwell's views are a fusion of neo-structuralism and theories of cognitive perception. It is interesting however to note that Bordwell himself downplays the importance of actual filmic information (music, images, dialogue) to the interpretive process and the construction of the narrative. Bordwell claims that "Interpreting (reading) is dissective, free of the text's temporality, and symbolic; it relies upon propositional language."[9]

---

[8] Seymour Chatman, Story and Discourse, Cornell University Press, 1978. While all the concepts in the book are discussed with regards to literary issues, Chatman tends to skip

Bordwell identifies three different structural elements in any narrative: *fabula*, *syuzhet* and style. The *fabula* and the *syuzhet* have been discussed above, and according to Bordwell are independent of the medium. The style "simply names the systematic use of cinematic devices,"[10] that is, the technical or material aspects of the specific medium. Bordwell acknowledges that the distinction between *syuzhet* and style is sometimes hard to make, and serves analytic purposes more than practical ones, but he insists that the "*syuzhet* embodies the film as a dramaturgical process, while style embodies it as a technical one."[11] Moreover, Bordwell asserts that:

> Film technique is customarily used to perform *syuzhet* tasks -- providing information, cueing hypotheses, and so forth. In the

---

on film examples to the literary concepts he cannot find filmic equivalences to. Thus, his book is useful in discussing literature, but is a weak analysis of film narration.

[9] David Bordwell, op. cit. p.30.

[10] Ibid. p.50.

[11] Ibid.

"normal" film, that is, the *syuzhet* system controls the stylistic system -- in Formalist terms, the *syuzhet* is the dominant.[12]

I believe that Bordwell is wrong in assuming the domination (or even distinction) of *syuzhet* over style. In fact, I believe that the *syuzhet* is style, technique and structure, and that these material aspects of film are the dramaturgical mechanism. The story elements are communicated through particular choices of words, images, sounds, points of views, etc. And as Bordwell himself acknowledges, these stylistic elements cue the perceiver toward making hypotheses and constructing the *fabula*. The problem with Bordwell's idea of separating style from syuzhet is that it enables him to discuss stylistic elements only when he wants to, and ignore them when he wants to assert a psychological, or other kind of explanation (see discussion of the "bull's eye" schema in chapter 3). While his analysis of perception is important and enlightening, the one of narrative structure is lacking.

Edward Branigan on the other hand is a structuralist of the cognitive narratology era. Not a postmodernist, Branigan is still constantly aware of the gap between the signifier and the signified, even (or particularly) in visual systems. Branigan writes:

> Representation is, as Umberto Eco suggests, "everything that can be used in order to lie." The resultant split of signified and referent is the basis of an indirect theory of experience where a viewing

---

[12] Ibid. p.52.

> encounters not the world but a system of codes and an ideology which bear a more complicated relation to the world.[13]

Branigan too is interested in perception, and how the viewer of a film constructs the narrative. But he sees the filmic text as a structural system cueing the perceiver toward the creation of meaning. Branigan's careful analysis identifies characters as just another structural element (like Barthes' agents or Greimas' *actants*), and thus dialogue, point of view shots, editing, etc. are all structural devices advancing the narrative by cueing the perceiver. While Branigan's analysis is very detailed, and much more careful in accounting for the different kind of material information provided by the film, he too stops short of a full cognitive account of perception. In my analysis I will mostly follow Branigan's lead, but I will also call attention to some of the limitations and oversights of his theory of point of view. I shall now turn to a close look at focalization, enunciation and their application to film discussion.

Focalization

In discussing cinematic point of view all film theorists refer to the literary concepts of narration and focalization. Mieke Bal defines focalization as "the relation between the vision and that which is 'seen,' perceived."[14] Focalization then designates a perspective, either physical or epistemic. There has been some confusion in literary narratology over the question of narration and focalization, where the two terms have been conflated to signify one thing, so that focalization assumed to dictate narratorial power as well. Genette tries to clarify the issue by saying:

---

[13] Edward Branigan, op. cit. p.183.

> . . . the theoretical works on this subject. . . suffer from a regrettable confusion between what I call here *mood* and *voice,* a confusion between the question *who is the character whose point of view orients the narrative perspective?* and the very different question *who is the narrator?* – or, more simply, the question *who sees?* and the question *who speaks?* [15]

Following Genette, both Bal and Rimmon-Kenan assert that focalization and narration are distinct activities; the narrator may be an adult recounting her childhood memories, while the focalizer may be the child experiencing the memory being told. It is, of course, possible (and even common) for focalization and narration to be combined and performed by a single agent.

Focalization has been often equated to point of view, or physical perspective. A distinction is made between an internal focalizer, a character, or an agent within the *fabula,* and external focalization which is done by an external narrator, functioning outside the *fabula*. Rimmon-Kenan suggests that "translated into spatial terms the external/internal

---

[14] Mieke Bal, Narratology: Introduction to the Theory of Narrative, University of Toronto Press, 1985, p.100.

position of the focalizer takes the form of a bird's eye view versus that of a limited observer."[16] This notion has been translated quite literally to film theory, where one assumes an omniscient, external narrator when the shot is all encompassing, and a character's point of view, when the physical perspective is limited. It is useful here to recall the work of Boris Uspensky who defined point of view not solely as a physical perspective, but as an expression of emotional, psychological and ideological states of mind.[17] Under this theory, a bird's eye shot in *Vertigo* (Hitchcock, 1958) can signify Scottie's psychological point of view (his fear of heights), while a physically restricted perspective in the same film may signify the stability of the external narrator. It is also important to mention that film's use of point of view structure (as a form of focalization) is generally reserved to the main character only, and is rarely and with great difficulty applied to other characters.[18]

Moreover, as George Wilson shows,

> First person narration is equated with film segments whose visual contents are meant to represent a character's visual experience. Most often, the shots in such a segment portray *what* the relevant character sees of his or her environment and, more or less, *how* he

---

[15] Gerard Genette, op. cit. p.186.

[16] Shlomith Rimmon-Kenan, Narrative Fiction: Contemporary Poetics, Routledge, 1983, p.77.

[17] Boris Uspensky, A Poetics of Composition, University of California Press, 1973.

> or she is supposed to have seen it. Alternatively, the shots may show certain objects and events as they are remembered, imagined, dreamed, and so forth by that character. Third person narration then subsumes all film narration that is not tied directly to the subjectivity of a character in these ways.[19]

Focalization here is therefore associated with first person narration, one that is embedded in a larger frame of narration (third person), one that is objective, or at least not tied to any character's subjectivity. But one needs to be cautious about equating focalization with subjectivity and third person narration with omniscience.

Focalization is a central and essential tool of narration; it may shift from one character to others and to the narrator(s), and these shifting relationships skirt around subjectivity in interesting ways. The shifts are significant for the construction of the *fabula* as each level of focalization subordinates the perceiver's knowledge to that of the focalizer. Focalization thus is a narratorial device that enables the narrative to withhold or release information as is needed for reasons of suspense, dramatic conflict, etc. The relations between the levels of focalization deserve much attention, and while a text may construct complex relations

---

[18] For a comprehensive account of cinematic P.O.V. see Edward Branigan, Point of View in the Cineam: A Theory of Narration and Subjectivity in Classical Film, op. cit.

[19] George Wilson, Narration in Light: Studies in Cinematic Point of View, Baltimore and London: Johns Hopkins University Press, 1986, p. 127.

between different character-focalizers, when it comes to external focalization Mieke Bal claims that:

> When external focalization seems to yield focalization to a character-focalizer, what is really happening is that the vision of the character-focalizer is being given within the all-encompassing vision of the external focalizer.[20]

This notion of embedded dominance where the external focalizer has more authority than the character focalizer, has also been adopted to film without careful examination. The general camera work is considered to be a kind of external focalizer (an effaced narrator or implied filmmaker), and it is considered to yield focalization to a character (mostly the main character). While the issue of levels of narratorial dominance will be addressed later too, it is important to mention that one does not need to accept Bal's position, as alternatives in the form of dialogic relations between different levels of the text have been offered by theorists like Mikhail Bakhtin.[21]

Enunciation

The concept of enunciation was first proposed by linguist and philosopher Emile Benveniste. Benveniste draws a distinction between *énoncé* and *énonciation,* the uttered and the enunciation. The utterance refers to words, sentences, the verbal text. The enunciation refers to the entire process of communication, and according to Bordwell it includes the act, context and the linguistic forms.[22] Benveniste drew another crucial distinction, the one between *discours* and *histoire*.

---

[20] Mieke Bal, op. cit. p.111.

> In *discours* the relation speaker-addressee/hearer is present (represented through, say, first and second person pronouns) but in *histoire* it is absent. The two styles, as conceived by Benveniste, are exclusive of one another.[23]

Histoire then, according to Benveniste, is a text that has no linguistic shifters indicating a source, a narrator. It is interesting to note that unlike most narratologists, Benveniste, in this sense, is an empiricist who is not looking for the ever-present narrator, but is willing to accept direct communication without mediation. A *discours,* on the other hand, is a text that bears the marks of enunciation, so the speaker inscribes herself into the text. For Benveniste then, enunciation and *discours* are exhibited in texts that employ overt narration, therefore leaving marks of subjectivity in the text.

Bordwell mentions that Benveniste's discussion of enunciation and the relations between *discours* and *histoire* is only a small part of his

---

[21] Mikahil Bakhtin, The Dialogic Imagination: Four Essays, Ed. Michael Holquist; Trans. By Caryl Emerson & Michael Holquist, Austin: University of Texas Press, 1981.

[22] David Bordwell, op. cit. p.21.

[23] Edward Branigan, op. cit. p.171.

writing, and a relatively undeveloped one.[24] Other theorists like Todorov and Genette have developed these concepts further, but have also made them their own. Genette, for instance, implies that any narration leaves traces of enunciation in the narrative discourse that has been produced.[25] Branigan mentions that Todorov interprets *discours* and *histoire* to "co-exist as aspects of a single phenomenon (which he terms *énonciation*). In this way he can say that the narrator of *discours* is still present in *histoire* though now *effaced*."[26] Todorov and Genette then fall into the category of rationalist theorists, ones who insist on the presence of a narrator, even when the narrator leaves no traces.[27]

Classical Hollywood filmmaking masks its own production, and presents itself as un-mediated representation (as if the viewer just "happens" to be an observer, a voyeur). Film semiotician Christian Metz claims that the camera is the "voice" of the filmmaker (or the narrator), and that the shots are acts of enunciation, so the film is actually discursive. But Metz claims that this discursiveness is masqueraded as *histoire*, as the film effaces all marks of enunciation.[28] Here, Metz falls into the category of the rationalists, in an attempt to posit a unified subject, an enunciator, behind the camera. A rationalist theory also constructs an entity on the side of reading, such as an implied reader, ideal reader, etc. The viewer, as far as Metz is concerned, shares the power with the camera/narrator, and even gets the feeling that she is the enunciator. We can clearly see that Metz is attributing to the camera and the viewer a power that Benveniste only attributed to the verbal activity of individuals within the diegesis. To a certain degree the camera is able to indicate space, but it does not automatically (or inherently) provide the deictics of "who" and "when." That is, even a physical view point shot can be understood to represent a character's perspective only in the larger context of the scene, i.e., through editing, sound and other devices. Nonetheless, shots that are associated with an omniscient narrator cannot embody those necessary shifters, and therefore cannot be considered to be enunciation in Benveniste's terms, although they fit under Genette or Todorov's descriptions.

---

[24] Unfortunately, the English translation of Benveniste's Problems in General Linguistics does not include the discussion of *énonciation, discours* and *historie* , which appeared in Vol. 2 of the original French version, so I can only rely on secondary sources.

[25] Gerard Genette, op. cit. pp. 212-214.

[26] Edward Branigan, op. cit. p. 171.

[27] The discussion of empiricism versus rationalism is based on Branigan's discussion of metatheory. See Edward Branigan, op. cit., chapter 8, pp. 168-171.

Edward Branigan claims that the relations between enunciation and enounced are analogous to the relations between narration and narrative.[29] Enunciation and narration are "that subsystem which implicates a subject in an activity: telling, watching, listening. The result

---

[28] Christian Metz, The Imaginary Signifier, trans. Celia Britton et al. Indiana University Press, 1981. pp. 91-96.

[29] Edward Branigan, op. cit. p.2.

of that activity is an object: *what* is spoken of, told, watched, listened to."[30] Earlier I have shown that narration and focalization are two distinct activities, which need not be collapsed into one procedure. The conclusion from these literary theories is that enunciation and focalization are distinct activities as well. While I am not sure this distinction would always hold true even in literary cases, I find it particularly problematic when discussing film. A distinction between enunciation and focalization will posit enunciation (and possibly even narration) as a verbal activity and focalization as a visual one. But surely, such a crude binary categorization cannot do justice to the multitude of irreducible sensory categories of film, such as vision, language, music, etc. Focalization and enunciation in film can each be exhibited both visually and verbally, and therefore the literary dialectics of showing/telling, or focalization/enunciation are inappropriate and provide a simplistic form of analysis. In the next section I exemplify the problem by a discussion of filmic point of view and verbal narration, both as enunciative and as focalizing acts.

Cinematic enunciation, focalization and point of view

Film is a visual and aural medium; it places the perceiver in certain spatial relations to the events it portrays. Following Branigan we may say that:

---

[30] Ibid.

> Film begins with space (for the camera must be some*where*) and works to create its time, while literature begins with time (for the verb must be inflected) and works to create its space.[31]

Film then, from its very first shot, places the perceiver in certain spatial relations to the events on the screen, and then works toward establishing temporal order, and action causality. The relationship of the perceiver to images needs to be discussed, as point of view and focalization are relevant not only within the diegetic world, but also on the level of the positioning of the perceiver with respect to knowledge. The relations between the perceiver and the filmic images are particularly interesting since the film posits the spectator as an invisible observer, a voyeur, which in classical cinema (unlike literature) is never directly addressed. The image is also produced as neutral or naturalistic, so the perceiver is rarely placed in a position to question his/her relation to the image. Before discussing the perceiver's point of view, I need to explain how this "invisible naturalness" is manufactured.

The cinematic (photographic) image is a representation of a pro-filmic event, one that took place in front of a camera, and was recorded by the photographer on film. The photographic image is the ultimate end of the mimetic tradition, as it provides a representation that seems identical to the pro-filmic event. In semiotic terminology we can say that the photographic image exhibits a strong bond between the signifier and the signified, and that this bond is both iconic (by resemblance) and indexical (by pointing to a pre-existing moment which was recorded).[32] The mimetic quality of the image led early film theorists to accept the image as an accurate and automatic representation of reality. Andre Bazin, for instance, hailed the realistic quality of film and encouraged filmmaking practices that resemble as much as possible the human visual perception. Thus Bazin called for long takes, the use of depth of field, and less editing. He praised the Neorealists use of on-location cinematography and non-professional actors. In this attempt to foreground "natural" human perception there is an embedded assumption that visual data is accurate to the real world and factual. Colin MacCabe claims that classical cinema constructs the image as primary, thus setting "an opposition between spoken discourses that may be mistaken and a visual discourse that guarantees truth -- which reveals all."[33] Since realism is not the main topic of this work, I will not elaborate on this fascinating problem. I will just say that like MacCabe, Branigan, Barthes and many others, I believe that film is a representational system, which creates an "impression of reality" by using cultural codes and systems of signification. "The real in this sense is the labor through which matter is transformed into significance by a spectator/artist."[34]

---

[31] From a correspondence between Edward Branigan and myself, as part of a discussion on the nature of the verb structure in the filmic image. June 1996.

[32] For a fuller discussion of semiotics, see chapter 3.

According to the realist approach the film constructs the spectator as an observer in the scene, placed in the best location, or as Todorov claims, when editing is accounted for, "an observer ideally mobile in space and time."[35] But being able to visually experience events of one scene from many different perspectives does not reflect the experience of a participant or an observer of the scene at a real time, and from an actual, specific point of view. Editing creates an "ideal observer," one that has omnipotence over the scene, while at the same time, being an invisible voyeur, she is occupying a ghostly role. Bordwell writes:

---

[33] Colin MacCabe, "Theory and Film: Principles of Realism and Pleasure" in Mast, Cohen & Braudy (eds.) Film Theory and Criticism, Oxford University Press, 1992, p.82.

[34] Edward Branigan, Point of View in the Cinema, op. cit. p.207.

[35] In David Bordwell, op. cit. p.9.

> It is not hard to find empirical fault with the invisible-observer account. It must ignore many stylized techniques which cannot correspond to optical processes (split screen, wipes, negative filming, "impossible" camera positions and movements). It presupposes continuity cutting to be the closest representation of actual perception. It forgets that even in ordinary films, the camera's position changes in ways that cannot be attributed to a shift in a spectator's attention. The model works wholly at a localized, "atomic" level: it seeks to explain only this cut or that image, not whole sequences or films.[36]

It is clear now that the relation of the perceiver to the film is much more complicated that just occupying the role of an "ideal observer." The text constructs an "impression of reality" which the spectator consumes and processes through a set of spatial and temporal positioning, which in turn serves as cues to the construction of the diegetic world.

But the relations between the perceiver and the diegetic world is not one of a binary opposition. According to Todorov, the diegetic world can be seen as a series of embedded levels or frames, where one part of the language contains or restricts another part.[37] An example would be a sequence which is narrated both by a character who participates in the diegesis and by an omniscient narrator who criticizes the character's judgment. This narrator is then mocked by the implied author, which enables the character to actually succeed because of the judgment they made. In such a text, the character's discourse is subordinated to the narrator's, which in turn is contained by the author's decision to grant the character respect. Embedding, or framing, then, implies a limitation on the epistemological status of each level of narration. While discussing embedding, Branigan shows how each embedded level functions as an object to the next level which "sees itself" as a subject, but is an object to the higher up level.[38] The characters then are objects to the subject narrator, who is, in turn an object to the subject spectator. Elsewhere, while discussing subjectivity, Branigan claims:

---

[36] Ibid. p.10.

[37] In Edward Branigan, op. cit. p.172.

> Subjectivity will refer to the narration given by a character in the narrative, but it should be remembered that, in actuality, each successive level of narration implicates a new subject -- a fictional or hypothetical perceiver -- in an activity of seeing (e.g. listening, telling, displaying) an object (i.e., what is seen, heard, displayed). 'Subject' and 'Object' are not fixed terms but indicative of a relationship between two elements.[39]

The largest frame within the text, the overriding omniscience, (the "implied author" or "image of the narrator") ends up becoming the object for the implied or real perceiver, the subject. Branigan concludes:

> The viewer's attitude becomes a composite of various hypothetical observers, characters, narrators, implied narrators, and the author.[40]

---

[38] Edward Branigan, Narrative Comprehension and Film, Routledge, 1992. pp. 66-67.

[39] Edward Branigan, Point of View in the Cinema, op. cit. p 2.

[40] Ibid. p.8.

Due to reception-based and cognitive-based theories it is easy today to see the perceiver as a subject interacting with the text, but before structuralism (and poststructuralism) it was impossible to regard the fictional character as an object at all. Realistic (and to a degree, too, modernist) fiction has treated the character as a subject and an individual. The character, after all, is responsible for the action (or at least part of it, in cases of natural or godly disruption). The character initiates, or responds to the dramatic conflict, and thus is the major force in moving the plot forward, towards its resolution. Rimmon-Kenan writes:

> The so called "realistic" argument sees characters as imitations of people and tends to treat them -- with greater or lesser sophistication -- as if they were our neighbors or friends, whilst also abstracting them from the verbal texture of the work under consideration. Such an approach. . . tends to speculate about characters' unconscious motivations and even constructs for them a past and future beyond what is specified in the text.[41]

The structuralist project has done much to demystify notions of the character as an individual which is larger than the sum of the textual references to it. For the structuralist, all signification can be found within the text, by deconstructing its structure, and by a semiotic analysis of its sign systems. Branigan asserts:

> Character is a construction of the text, not *a priori* and autonomous. It is not a "first fact" for literary criticism through which the remainder of a text is interpreted, made intelligible. Rather,

---

[41] Shlomith Rimmon-Kenan, op. cit. p.32.

> character exists to serve and to mask unconscious forces as they are played out in a drama which implicates the viewer.[42]

Moreover, Benveniste shows us that the pronoun "I" designates different persons in the duration of the discourse, as the "you" becomes a new "I." In other words, during a conversation "I" would alternately designate the different participants in the conversation, as they express their own opinions. For Benveniste, language is the site of subjectivity, as the ego exists only when contrasted with an "other," expressed through a pronoun.[43] Roland Barthes, too, claims that:

> Linguistically, the author is never more than the instance writing, just as *I* is nothing than the instance of saying *I:* language knows a "subject" not a "person," and this subject, empty outside of the very enunciation which defines it, suffices to make language "hold together," suffices, that is to say, to exhaust it.[44]

For this exact reason enunciative shifters are important as markers of traces of subjectivity in the text. But this subjectivity does not refer to existents, but rather to discursive features of the text.

Benveniste's and Barthes' linguistic analysis supports the structuralist effort to deplete the notion of the character from any excess meaning, and serves as a constant reminder of the literary device, i.e., verbal language. But this linguistic and structuralist minimization of the character was not easily received by everyone. Seymour Chatman writes:

---

[42] Edward Branigan, op. cit. p.12.

[43] Emile Benveniste, "Subjectivity in Language" in Problems in General Linguistics, trans. by Mary Elizabeth Meek, University of Miami Press, 1971 (original French version was published in 1966). pp. 223-230.

[44] Roland Barthes, "The Death of the Author" in David Lodge (ed.) Modern Criticism and Theory, Longamn, 1988, p.169.

> The equation of character with "mere words" is wrong on other grounds. Too many mimes, too many captionless silent films, too many ballets have shown the folly of such a restriction. Too often do we recall fictional characters vividly, yet not a single word of the text in which they came alive; indeed, I venture to say that readers generally remember characters this way.[45]

Even a theorist like Bordwell, who places the critical weight on the viewer, is not completely free of romantic notions of the character as a subject, and not just a structural object of narration. Bordwell attributes all technical properties of the film as deriving out of the need to support the character's "transmission of *fabula* information, with the result that bodies and faces become the focal points of attention."[46] Such an attitude leads toward a limited and restrictive notion of point of view, by which Bordwell recognizes only a subjective, and mostly physical, view-point of a character.[47] In other words, for Bordwell, point of view shots are one of many techniques used to assert the subjectivity of a character, as the center and major force of the *fabula*. If the point of view shot cannot be justified by physical perspective, it is justified as a psychological one, thus further asserting the character as a subject, one with conscious and unconscious needs and desires.

In contrast, Branigan points out that physical view point shots are not enough to indicate subjectivity. As an example we can look at a film like *Lady in the Lake* (Montgomery, 1946), which is basically a single traveling point of view shot, but critics agree that it fails as a subjective film. Branigan claims that mere perspectival POV shot is not necessarily analogous to the experience of being that character, or feeling that character's feelings.[48] For him,

---

[45] Seymour Chatman, op. cit. p.118.

[46] David Bordwell, op. cit. p.162.

[47] Ibid. See discussions of *Lady Windermire's Fan* (p. 178-186) and *The New Babilon* (p. 257-260).

> What is important is not the camera as an absolute reference point but the relation among camera, character, object and a perceiver's hypothesis about this relation.[49]

While I accept Branigan's position, I would also like to add that for a point of view shot to work (i.e., to convey the epistemological status of a character), a relation between the POV shot and other shots, sound design, and editing needs to be accounted for. Branigan spends much of his book discussing different constructions of POV shots (or actually, as he rightly points, a pair of shots whereby we see a glance and an object being looked at), but he refrains from discussing the functioning of this pair within larger material contexts. The filmic medium struggles with conveying psychological states of characters. Unless voice over narration is used, there is no direct verbal access to characters' thoughts, except for what they say. To compensate for that, the film needs to show the character's reaction to the events, so that the viewer can infer beyond the verbal text. The pair glance/object in the POV shot is useless without a third shot of the character conveying the reaction to what has been seen. Granted, this reaction can be conveyed already in the first shot, that of the glance, but then the perceiver is unaware of what the character is seeing, and may therefore be unable to interpret the reaction. But understanding the importance of a knowledge gain (on the side of the character) is not conveyed solely by the response (body language or voice) of the reaction shot. Music, sound effects and editing, all cue the perceiver toward understanding what the character knows and how she feels about it. I would therefore suggest a small correction of Branigan's statement, whereby POV functions as a relation between all textual information (including the camera), character, object and the perceiver.

---

[48] Edward Branigan, op. cit. p.7.

[49] Ibid. p. 110.

For Branigan:

> Point of view as a system of the text functions to control (expand, restrict, change) the viewer's access, not to a real object (through a "camera"), nor to psychological states and attitudes, but to signification. *Points of view are epistemological boundaries inscribed within the text.*[50]

This view sits well with a theory of embedding. Branigan's notion of point of view, much like literary theories of levels of enunciation and discourse, functions as a structural theory, whereby point of view becomes a device in creating a level of narration, which can then be affirmed, denied, or restricted in another level of narration. Branigan asserts:

> Even so, point of view is only a partial description of the movement of narration through a text. For example, although the choice of point of view is frequently analyzed in terms of the information which is thereby suppressed, other effects -- such as the overall management and delay of enigmas, the arrangement of action sequences, and soliciting the reader's, or spectator's interest -- are

---

[50] Edward Branigan, op. cit. p.178 (my emphasis).

> more global in nature and are best analyzed when narration is considered as an interlocking system of many levels.[51]

For Branigan, then, the literary distinction between focalization and narration does not hold true in the case of film. The two are not distinct and separate activities: focalization is one of many narratorial devices.

Like Branigan, I too, believe that in the case of film, enunciation and focalization are not distinct activities, but are effective narratorial instruments. In the next paragraphs I will give a few examples from research done in the past few years.

While most films do not employ any form of overt verbal narration (other than dialogue), a small group does use voice-over narration, as a principal mode of transmission of the narrative. It may seem that this group exhibits narration as a separate activity from focalization, but as we shall see, this is not so.

David Black performs an adaptation of Genette's theory of levels of narration to cinema. According to Genette, the fundamental level of the narrating act (the one that initiates the discourse), done by a narrator which does not participate in the events, is the extradiegetic level. When a character narrates, she does so in a diegetic or intradiegetic level, and what is being recounted by the character exists at the metadiegetic level. But who is the extradiegetic narrator in film? While trying to locate that function, Black proposes that the film material (sound and image) is the fundamental narrator, and he proposes to call it an intrinsic narrator, since it refers to the discursive activity of the medium itself.[52] Black justifies this distinction by claiming that verbal narration in film is a product within a larger system of images and sounds, that is, it is not the initiator of the discourse, in the same way that the literary extradiegetic narrator is. Black claims:

---

[51] Edward Branigan, Narrative Comprehension and Film, op. cit. p. 115.

[52] David Allan Black, "Genette and Film: Narrative Level in the Fiction Cinema," Wide Angle, 8 (1986): 19-26.

> Even voice-over narrations that initiate the language and enjoy the spatio-temporal abstraction of the novelistic extradiegetic narrator are, nonetheless, entities of a secondary fictional order, included within an enveloping discourse.[53]

In support of this thesis we can look at Sarah Kozloff's comprehensive work on voice over in cinema. Kozloff shows that omnipresent, external, third-person narration is rare in cinema, and that most voice over narration is done in first person by a character who takes part in the events.[54] Moreover, Kozloff shows that unlike literary narration, filmic voice over narration is intermittent and interwoven with dramatic scenes. The voice-over narrator then is not the primary source of discourse (Genette's extradiegetic narrator), but one of few narrators (a homodiegetic one). Like Black, Kozloff concludes that

> A homodiegetic voice-over narrator is always subsumed by, and thus subordinate to a more powerful narrating agent, the image maker who dramatizes the story on the screen.[55]

We can see now that even films that use voice-over narration, do so not as an initial level of narration, but as an intermediate one, an embedded level, which may (or may not) have more epistemological authority than the characters in the film, but is definitely subordinated to a more basic level of narration, that of the intrinsic film medium. It is interesting to recall in this context Colin MacCabe's assertion that film posits voice as epistemically less reliable than image. While MacCabe may support his claim by historical, or semiotic explanation, it is clear now that the status of the voice as subordinate is also inherent to the structural restraint of film narration, i.e., to the fact that the fundamental level of film narration is not solely verbal. Voice-over films, much like the so called un-mediated films, use enunciation side by side with focalization toward purposes of narration.

---

[53] Ibid. p.20.

[54] Sarah Kozloff, Invisible Storytellers: Voice-Over Narration in American Fiction Film, University of California Press, 1988, chap. 3 & 4.

[55] Ibid. pp 48-49.

Finally, it is worth mentioning that film focalization (via a point of view structure) can function in much the same way direct discourse does. Jeffrey Rush performs an adaptation to film theory, of the quotation principles of direct, indirect and free indirect.[56] While Chatman struggles with addressing forms of quotation, character thoughts, and interior monologue in film,[57] Rush finds visual analogous to all three phenomenon. Rush claims:

> The quotation marks, the disowning implicit in direct discourse, have analogs in the classical point of view film sequence. In a point of view sequence, the extradiegetic narrator appears to vanish. The camera seems motivated entirely by the character's glance and, through that, by the character's thoughts themselves.[58]

---

[56] Jeffrey S. Rush, "Lyric Oneness: The Free Syntactical Indirect and the Boundary Between Narrative and narration," Wide Angle 8 (1986): 27-33.

[57] See Seymour Chatman, op. cit. Chapter 4. While a large portion of the chapter discusses narration and speech acts (pp 161-195), only the last couple pages directly address some of these issues with regards to film. The discussion however, is problematic and insufficient.

[58] Jeffrey S. Rush, op. cit. p.28.

As an example Rush recounts the sequence from *Psycho* (Hitchcock, 1960), where Marion is leaving town, and she drives by her boss. We see the boss' puzzlement from Marion's perspective, through the mirror, and then the camera cuts to show Marion as she gapes and hesitates.

> Her thoughts are as clear and as immediate as they would be in a sentence, " 'Oh not him of all people,' thought Marion as her boss passed in front of her. 'Should I turn back now?' "[59]

In indirect discourse the narrator is responsible for summarizing the thoughts of the character, so the quotation marks are removed, and the speech is being interpreted, rather than imitated. A filmic example can be seen in *The Butcher* (Claude Chabrol, 1969) where a teacher feels lonely and isolated from her physical environment. The camera shows her at the school window and then pulls back to show the entire stone face of the schoolhouse with the woman, small in one window. This kind of shot conveys the woman's feeling in a metaphoric way, by creating an image of seclusion and loneliness. The intrinsic narrator, (in this case the camera, rather than editing or another device) is responsible for conveying how the woman feels, without cutting to a physical POV shot (or direct discourse). The analog of free indirect discourse (one that omits the direct attribution) can be equated to Uspenski's psychological or emotional point of view. A film sequence can convey a character's mood, or the narrator's judgment about a character's mood, without explicitly showing the character experiencing that mood. Rush's example is a sequence from *Eclipse* (Antonioni, 1962), where shots of empty space (empty of characters, that is), convey Vittoria's need to claim a place for herself.

---

[59] Ibid.

Appropriating Rush, we may say that in certain cases a film's focalization (whether done by a character or a narrator) is the equivalent of enunciation. Under this view, visual constructs can convey a subjectivity and at the same time they presume the voice of the narrator. Once again, we see that enunciation and focalization are devices in narration, and that in film, they often are interchangeable or collapsed into one activity.

Summary

Part of the problem with the attempt for a clear-cut distinction between enunciation and focalization is that film communicates stories via sound (verbal and non-verbal) and image. Early film theorists attempted to understand image in verbal terms and ignored the complexity which is the result of the multiplicity of sensory channels. In his effort to define film as a language, or at least a language system, Christian Metz says that

> A close-up of a revolver does not mean "revolver" . . . but at the very least, and without speaking of connotations, it signifies "Here is a revolver! . . . Even when the shot is a "word," it remains a kind of a "sentence-word," such as one finds in certain languages.[60]

For Metz, we think and understand photographs in terms of words. Many objections have been proposed to Metz's claim, like the issue of negation, or positive assertions, and the fact that when the image is not a close-up of one object, there are many alternative accurate descriptions of the content and meaning of that picture. In responding to Metz's assertion Branigan writes:

---

[60] Christian Metz, Film Language: A Semiotics of Cinema, trans. by Michael Taylor, Oxford University Press, 1974, p. 67.

> The possibility that both diegetic and mimetic processes may be involved in processing a picture suggests that the interpretive consequences of a picture may be greater than either the pictorial associations of the picture seen or the words we apply.[61]

Branigan is trying to open the door for an interpretation which is both verbal and non-verbal, but his concept of pictorial interpretation is elusive and vague.

Metz is also not alone in interpreting images in verbal terms. In fact, many theorists assume, like Chatman, that "since cognition is already a verbal constitute, or is easily reduced to one, its transference to verbal narrative is simple and immediate."[62] Chatman creates a subordinate system by which visual perception has to be understood in verbal terms, since cognition is verbal. But as will be shown in chapter 2, recent cognitive theories of visual information processing suggest that images are not necessarily understood in verbal terms, and they give accounts of visual cognition that is independent of language cognition. The merit of these theories is in enabling us to think of distinct narratorial computation systems for words and pictures. Under this approach we will have to look at enunciation and point of view as they function visually, verbally (and maybe in non-language based aural system as well), toward the construction of visual, verbal, and (other sound) narratives. A multiplicity of narratives (or narrative levels) in one text will be studied then, in a complex model that accounts for the material complexity of film as a medium. And after all, since psychoanalysis long gave up on the idea of a unified subject, there is no need to fear the idea of a complex and sometimes contradictory narrative text. In the next chapter I will look at the differences between the perception and cognition of images and natural languages. In following chapters I will apply a cognitive approach to film narration and comprehension, and particularly to the issues of point of view and filmic enunciation.

---

[61] Edward Branigan, "Here is a Picture of No Revolver; The Negation of Images, and Methods for Analyzing the Structure of Pictorial Statements," Wide Angle 8 (1986), p.11.

[62] Seymour Chatman, op. cit. pp. 181-182.

Chapter 2

Cinematic implications of cognitive science research on the perception and cognition of images and language.

Introduction

Film communicates simultaneously through images, dialogue, sound effects and music, and it is difficult to separate (both experientially and theoretically) the impact of each one of these communicative channels on our interpretation of the film. Action, camera movement, point of view editing, voice over, and other structural filmic devices, further complicate the discussion. In this chapter I will first discuss the concept of meaningful perception, or the relationship between perception and cognition. I will then describe the fundamental differences between the perception of images and language (as it is conveyed in the dialogue), and how this information is being cognitively processed as both propositional sets (language and images) and holistic images. In addition, I shall discuss memory storage and retrieval mechanisms, as these are crucial for the reorganization of syuzhet material into fabula (or the construction of the narrative). This chapter is meant as a general overview of research from cognitive science which is relevant to film comprehension. The application of the research outlined below to film narration and interpretation will be developed in chapters three and four.

While discussing perception I would like to use examples from Sergio Leone's 1969 epic Western *Once Upon a Time in the West*. This film starts with a twelve-minute credit sequence, establishing a solitary train station in the West, and three gunmen waiting for the train. In the entire scene there are two lines of dialogue, delivered by the scared station manager, who is locked up within the first three minutes of the film. The next nine minutes feature sound effects that are generated by the image (a fly, dripping water, wind, and finally the arrival of the train). No words or music accompany the scene. The sequence ends with the arrival of the train, and its almost immediate departure, leaving the men thinking that the man they were waiting for did not come. But as the train pulls away, we hear the sound of harmonica playing, and then see Charles Bronson playing it. A short dialogue reveals that the men are intending to kill Bronson, at the end of which they all pull guns and shoot at each other. Bronson kills all three men, and is wounded himself in the shoulder. At this moment, seventeen minutes after the beginning of the film, the plot starts, and is going to become progressively more complex for the next three hours. The scene is composed of mostly static camera shots, which are divided into some wide angle shots exposing the spatial relations, and the characters' placements with regards to one another, and many close-up shots of faces, or objects. The scene is long and slow, featuring much repetition of detail, and is communicated almost solely by visual means. The few sound effects, rather than expanding the boundaries of the frame (as in a sound that indicates the arrival of someone into the frame), are originated from within the visual field, and are always explained visually (like the squeaking noise whose source is indicated by a visual cut to a metal windmill). The lack of dialogue or action, and the static nature of the camera work, make this sequence useful for an analysis of visual perception and cognition. Other films that will be discussed in this chapter are *Alien* and *The Silence of the Lambs.*

Meaningful perception

Within three months of a child's birth, the child can see with considerable clarity the world around it.[1] Within the first year of a child's life, the child learns how to crawl and walk, and in so doing learns to make a connection between the visual representation in his/her brain and the physical world. As coordination develops, the child can avoid objects in its way, navigate around space, and reach and grasp with great accuracy. In order to perform all these tasks the child's visual perception needs to be in full operation. This operation precedes the use of language (which starts developing after a year of age) and thus precedes the ability to name objects and to define spatial relations verbally. Visual perception not only precedes language, but is also very difficult to describe in linguistic terms. Our visual perception is fast (compared to language perception), operates simultaneously on many objects in the visual field, seems automatic, and we have no direct awareness of how the experience of perception comes about.[2]

Unlike vision, the sense of hearing is already active while the fetus is inside the mother's womb. But the ability to hear sounds, to enjoy music, or be scared of a loud thunder is quite different from the acquisition of natural language skills. Language is a highly complex and codified communication system about the world. It uses arbitrary chains of sounds to represent objects in the world. It is abstract, and has come about by a lengthy social (and historical) process. When the child learns how to speak it first learns general nouns. More than half will refer to classes of concepts (like "dog," "cat," "house," etc.), and only between 18 months to two years will the child start putting together two word utterances.[3] By the age of four the vocabulary consists of 1500 words, and the use of grammatical forms approaches adult level. By the time children enter school they comprehend, on average, more than 14,000 words.[4]

---

[1] Mark H Johnson and John Morton Biology and Cognitive Development: The Case of Face Recognition, Oxford UK & Cambridge, USA, 1991, Ch. 2 pp. 23-37.

[2] Nicholas J. Wade & Michael Swanston, Visual Perception: An Introduction, London and new York: Routledge, 1991, pp. 1-5.

Language also varies across cultures, and a person landing in Baghdad with no knowledge of the Arabic language and no translator, will find it difficult to function in that society. But that same person will still be able to see Baghdad, will know better than to cross the road when a car is approaching, and will probably be even capable of hailing a cab, using mostly visual perceptual skills and body language conventions for cab-hailing. In order to understand language, we need to first learn it: we need to individually learn each and every symbol, and the referent it stands for, plus the social rules and conventions that govern language use. We also need to be able to store this lexicon and the set of rules in our memory, so that we can use it to decode the sounds that we hear. It is clear therefore that the perception of language becomes meaningful only when it is cognitively processed by high-order mechanisms such as memory retrieval, matching and sorting, and distinction into categories.[5]

---

[3] L.R. Gleitman & H. Gleitman, "A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition," Current Directions in Psychological Science, 1992:1 pp. 31-35.

[4] P.A. deVillers & J.G. deVillers, "Language Development" in M.H. Bronstein & M.H. Lamb (eds.) Developmental Psychology: An advanced textbook, 3rd ed. Hillside, NJ: Earlbaum, 1992, pp. 337-418.

[5] Adrian Akmajian, Richard A. Demers, Ann K. Farmer and Robert M. Harnish,

Images function differently. Not only does their perception seem automatic and fast, but they seem to possess, and not necessarily acquire, meaning: viewing a dog is a meaningful event even before we know what it is. If the dog is agitated and barks we may feel fear; we will classify it as a dangerous animal and respond by taking precaution. But the word "dog" is meaningless until we learn what it refers to, so the warning "watch out for the dog" will be meaningless until "the dog" is attached to a specific referent (a specific canine, who is capable of barking and biting, etc.). It seems to follow then that at some basic level visual perception functions by *just seeing* things, even when one does not know what one sees. Philosopher Fred Dretske makes a distinction between two kinds of perceptual processes:

> The perception of objects -- what I will call *sense perception* -- is that early phase of the perceptual process that culminates in sense experience (visual, auditory, tactile, etc.) of the object. Perception of facts about these objects, on the other hand -- that which constitutes *meaningful perception* -- is a more inclusive process. Besides sense perception, meaningful perception includes a knowledge (at least a judgment or belief) about the object being experienced.[6]

According to Dretske then, in order to have a meaningful perception, one needs to involve high-order cognitive activities such as retrieval

---

Linguistics: An Introduction to Language and Communication, Cambridge, MA: MIT Press, 3rd Ed. 1990, particularly chapter 6: "Semantics: The Study of Meaning and Denotation." See also George Lakoff, Women, Fire and Dangerous Things: What Categories Reveal About the Mind, Chicago, IL: University of Chicago Press, 1987, p.2 and chapter 2.

[6] Fred Dretske, "Meaningful perception," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995, p.331.

of previous knowledge, and mechanisms of judgment and belief about the object being perceived. Object recognition, which is based on categorization and memory, typically does (but need not) involve the retrieval of verbal labels (i.e., naming). Dretske asserts:

> Meaningful perception requires more than good eyesight. It requires the kind of conceptual skills needed to classify and sort perceptual objects into distinct categories.[7]

According to this approach then, meaningful visual object recognition requires two processes: discrimination and categorization. In discrimination one perceives the features of objects and other visual qualities. In categorization there is a process of recognition that is based on reliance on memory (at least of categorical information). Dretske does not specify the features of categorization, but in most cases categorization typically involves naming, and organization in propositional sets. A rare example in which meaningful perception can be executed without naming and propositional sets can be seen in a guide dog for the blind, who is capable of discrimination and will not cross the road when a fast car is approaching, but is incapable of naming the objects "car" or "road" altogether. Still, the dog's visual perception and discrimination skills have resulted in a meaningful perception, one based on belief, or judgement, which require pre-learned knowledge of certain categories. But the guide dog example also shows that many complex cognitive mechanisms operate in order to achieve ability to form a judgement ("safe to cross"). The dog needs to evaluate the situation, compare it with stored memories or learned sequences of behavior, and then make a judgement about it. Whether it requires language or not, then, meaningful perception for Dretske is a product of the application of complex conceptual skills, ones that have been learnt and cognitively processed before perception can become meaningful.

---

[7] Ibid.

The question of the stage at which perception becomes meaningful is particularly important in the case of film comprehension. Film comprehension occurs at two distinct temporal frames that are interrelated but not identical. As the film is being projected, images and sounds are being perceived and the actions and events that they portray are being interpreted as they relate to the plot. But the overall narrative of the film can be inferred and concluded only after the film has ended, and is therefore a post viewing product. In order to come up with this conclusive narrative text, the perceiver needs to retrieve filmic events from memory and re-organize them in a causal and temporal linear chain. Memory retrieval, problem solving operations, and re-organization of the filmic material are higher order cognitive activities, which are likely to necessitate the use of language, or language-like systems. But this post-viewing narratorial operation is dependent on what was stored in memory during viewing, and how that material was perceived and cognitively processed in order to be stored in memory. Now, if perception becomes meaningful only when categorization is involved, as Dretske claims, and since memory is a high-order cognitive activity, it is unlikely that sense perception will be categorized and memorized as such. Only *meaningful* perceptions will undergo cognitive operations and will be stored in memory. In other words, visual perception will first be processed and understood in some language-like way, and only then stored in memory. It then follows that the construction of the narrative as a high-order cognitive activity relies mostly (or maybe even exclusively) on language and language-like information systems for its operation. Yet some visual events that seemed non-meaningful at the time of film perception may become meaningful later on because of the needs of the evolving plot, and for the construction of the conclusive narrative (see discussion of *The Usual Suspects* in chapter 3). And if visual perceptions are only stored in memory once they become meaningful (i.e., categorical), then some non-meaningful perceptions may not be available for re-evaluation at the time they are needed for the construction of a comprehensive narrative. Dretske claims that

> Some form of constructivism or computationalism is *therefore* inevitable for meaningful perception, for seeing facts.[8]

But what if, like the guide dog, we do have mechanisms for meaningful visual perception during the viewing of a film, which involve judgements (like "safe to cross") yet do not involve such high-order cognitive activities as computation and categorization? And what if those meaningful visual perceptions are actually stored in memory for later use? It will then follow that our construction of the narrative is not solely dependent on language-based (propositional) stored information, but is also somehow affected by meaningful visual perception and cognition. In other words, at least in the case of visual perception, meaningful perception may operate at lower, less complex levels of cognitive activities, ones that could be stored in memory without being converted by computation into sets of propositions. It is interesting to note here that the cognitive literature on mental imagery and visual memories is engulfed in a similar debate about the nature of visual cognition, and how it is stored in memory for further use. The debate, referred to as a debate between the depictivists (or imagists) and the propositionalists, offers a complex picture, and at this point seems to show that both operations are active at the same time.[9] I would like to now to turn to account for the mechanisms of sense perception and how it becomes meaningful for vision and for language in the case of film.

---

[8] Ibid, p.344. My emphasis.

Bottom-up and Top-down Perception

Any discussion of perception, whether aural or visual, involves a discussion of low-order and bottom-up perception, and its relation to high-order cognitive mechanism, and to top-down operations. Low-order cognitive operations are those that guide our attention to discrimination of features in the visual field.[10] Higher order cognitive activities refer to inductive reasoning, memory activation, categorization, naming, problem solving, inferencing, and such operations which are performed on already perceived sensory information. Bottom-up perception refers to the actual consumption of sensory information from the world and its processing by the brain.

---

[9] For further discussion of the debate between the Depictivists and the Propositionalists see Stephen M. Kosslyn Image and Mind Cambridge, MA: MIT Press, 1980. See also P.N. Johnson-Laird, Mental Models. Boston: Harvard University Press, 1983, ch7.

[10] Attention and discrimination can be considered as high, or medium-order visual perception levels of cognition. In my use of the low-level term throughout this dissertation I do not try to challenge the cognitive science literature on the subject, but just to indicate that object recognition, and discrimination are quite different than high-order cognitive operations such as memory storage and retrieval, and organization of narrative information in causal order.

Top-down operations refer to devices like feedback mechanisms, in which memory and expectations effect low level processing, and determine attention patterns (see discussion on the influence of ideology on perception by the materialist branch of reception studies in chapter 4). In the case of film there is an a-priori top-down operation, namely the knowledge that we are watching an artistic product that has been filmed in the past, and is now projected on a screen. This top-down operation is of course historically acquired: children often have nightmares about Bambi's mother's death, being unable to recognize the fictional nature of the experience. It is interesting in this context to mention that Lumière's 1895 viewers -- the first film viewers ever -- were allegedly overwhelmed by the image of a train approaching the camera/audience, and they ran away from the theatre in fear. Given that they had not yet learned what the film experience is, they had no top-down beliefs at their disposal, and were therefore reacting to the bottom-up perception that a train is about to run them over. But for most adults of the first world at the turn of the millennium, there is a clear top-down comprehension of what film is about (as an artistic and as a sensual experience), a mechanism which encompasses or brackets the events that occur in the diegetic world of the film's plot.

But even within the narrative, as well as in our living experience, for bottom-up perception to be cognitively processed there usually needs to be a higher order cognitive operation involved. For instance, the second shot of *Once Upon a Time in the West* shows a slow tilt-up of the camera from the boots of a man, up his legs, his gun, his coat and finally his face under a felt cowboy hat. In order to recognize the objects "boots" and "gun" we need to have seen them once before (in real life or in film), learned what they are, and be able to name them. If we don't know what the gun is (an instrument that can kill), we won't understand the look of fear and surprise on the face of the station's manager in the following shot. To use Dretske's terminology, while sense perception will just be seeing the gun as a metal instrument of a certain shape, its meaningful perception will be to see it as a "gun," (or make a judgement that it is a gun) with all its usage and social implications. For meaningful perception to occur here, there needs to be a higher-order procedure in operation; a recognition of the object with reference to other ones seen before. Moreover, since much of the construction of the narrative activity occurs in retrospect, it heavily relies on higher order operations such as analysis and reorganization of stored memories. But as I claimed earlier, although meaningful visual perception may not require computation and propositional sets, higher order operations like memory and categorization certainly do.

It is particularly important to look at low level perception, since not all films rely solely on "meaningful" perception (like the "this is a gun" case described above). In some films, the question is "what are we seeing?" and the answer is not reduced to an object. In the first *Alien* (Ridley Scott, 1979), for instance, the Alien can never be seen as a discrete object. When first encountered it is barely lit, seen through mist and fog, usually through the masks the team is wearing, or through the video signal sent to the ship. Once on the ship, it is mostly shown in a series of close-up shots which reveal only parts of the alien. Throughout the film, it changes shape, form and material, and remains a concrete form of threat precisely because the audience cannot easily identify it, and therefore cannot make predictions about what it is capable of. Because the Alien is never categorized and defined, the audience needs to pay constant attention to low-order visual perception in dealing with its threat. Meaningful perception of the Alien, as far as Dretske would postulate, occurs only at the level of computing the responses of the other astronauts to the Alien, as well as in top-down expectations that what we see is the alien, or the consequences of its actions. Since the audience can never anticipate how and in what new form the Alien will present itself, and since the film refrains from actually showing the Alien much, the audience relies on information derived from other characters. *Alien* is an example of a film where low-order perception is crucial, but insufficient in and of itself for comprehension. I shall therefore carefully examine here bottom-up theories of low-level perception and how they affect meaningful perception. In following chapters I will show the importance of such low-level perception to higher order mechanisms of narrative comprehension.

Only a few researchers support a purely bottom-up approach to perception. One of the more influential ones is Gibson who in the 1960s re-focused attention on the direct theory of perception.[11] Under this position

> the stimulus (. . .) contains all the information needed to specify the distal state of affairs. If the proximal stimulus is understood, not as a static distribution of energy occurring on the receptor surfaces at a time, but as the total dynamic pattern

---

[11] J.J. Gibson, The Ecological Approach to Visual Perception. Boston, MA: Houghton Mifflin, 1979.

of stimulation reaching a mobile observer over time, there is no need for inference, reasoning, and problem solving.[12]

This approach eliminates the possibility of any cognitive activity during perception and assumes that perception functions simply by a faithful registration of stimulus information, and "the perceiver does not contribute anything to the act of perception [. . .]."[13] But Gibson's theory has been challenged quite often in past decades, both on theoretical and on experimental grounds.[14] It is clear today that in most experiences of perception there is some sort of interaction between high-order cognitive activities and low-order perceptual ones. As a result, many theorists started exploring the relationship of the "top-down" and the "bottom-up" mechanisms.

These investigations resulted in a range of concepts like Minsky's Frames[15], and Schank and Abelson's Scripts.[16] All these researchers were looking for an appropriate frame of reference, a context-based unit of knowledge which is used to process new information. According to these theories there is a hierarchy, or architectural structure for cognition, which means that new information is always matched and compared with older, previously compiled large-scale data structures. This kind of context driven architectural theory explains the power and speed of mental activities. Marvin Minsky notes that:

---

[12] Fred Dretske, op. cit. p.342.

[13] J.J. Gibson, "On Theories For Visual Space Perception," in G. Jansson, S.S. Bergström & W. Epstein (eds.) Perceiving Events and Objects, Hillsdale, NJ: Lawrence Erlbaum Associates, 1994, P.183.

[14] For an example see J. Fodor and Z.W. Pylyshyn, "How Direct is Visual perception? Some Reflections on Gibson's 'Ecological Approach'," Cognition 9, 1981.

[15] Marvin Minsky, "A Framework for Representing Knowledge" in P.H. Winston (ed.) The Psychology of Computer Vision, NY, McGraw Hill, 1975, 211-277.

[16] R.C. Schank and R.P. Abelson, Scripts, Plans, Goals and Understanding, Hillside, NJ: Earlbaum, 1977, Chap. 3.

> In popular culture, memory is seen as separate from the rest of thinking; but finding the right memory - it would be better to say: finding a *useful* memory - needs the same sorts of strategies used in other kinds of thinking![17]

Minsky is not alone in linking memory, thinking, and information processing. In fact, all the theories mentioned above are historically based, in the sense that they all focus on the retrieval mechanism of a context based unit. The theorists diverge however when discussing the nature of the cognitive architecture and the retrieval procedure, as will be shown later.

While all these theories focus on a cognitive structure which is already established (by past experiences and previous knowledge gathering) Schank and Abelson propose a useful model for discussion of the perceptual activity. They claim that different situations trigger different memories, experiences and appropriate contexts. Thus when one mentions a restaurant, a whole set of concepts emerge: restaurants as related to food, to social activity, as different from fast food chains, a particular restaurant one has recently been to, etc. Schank and Abelson call these context/memory-based units, "scripts:"

> [. . .] a script is a predetermined, stereotyped sequence of actions that defines a well-known situation. Scripts allow for new references to objects within them just as if these objects have

---

[17] Marvin Minsky, op. cit., 174.

> been previously mentioned; objects within a script may take 'the' without explicit introduction because the script itself has already implicitly introduced them.[18]

For Schank and Abelson the script is a very specific unit designed for a particular context, and is not subject to much change. The specific nature of the script creates a complex hierarchical structure in which many scripts have to be pulled at any given situation, the specific information has to be first matched or compared to all the relevant scripts and only then understood.

Schank and Abelson's description of the script is very similar to literary and filmic descriptions of a genre. The genre piece shares a limited variety of themes, plots and the same gallery of characters. The topic of the film, the kinds of events and developments, are chosen from a relatively narrow range of possibilities. A well-trained reader/viewer of a genre piece has a rigid set of expectations about the text, before even starting to perceive the text itself. The Western consumer, for instance, will expect to see

> (. . .) a succession of muddled brawls in bar-rooms, tense and inscrutable poker games in smoky saloons, gunfights in empty streets, showdowns among the rocks with whining bullets, cavalry pursuits and Indian ambushes, mysterious strangers riding into town in search of vengeance or redemption . . .[19]

Like the script application, a viewing of a new Western is always compared and processed through the knowledge of the genre, and its predetermined scripts. In *Once Upon a Time in the West*, for instance, the three gunmen wear dark clothes and black hats. Charles Bronson, on the other hand, wears a white coat and a white hat. This color code indicates to the Western connoisseur (and to some degree to the film competent viewer of other genres as well) that Bronson is probably the good guy, while the other three are likely to be bad. When minutes later Bronson successfully kills all three, ties his wounded arm to his coat and gets up to leave, the audience is barely surprised, even though they just spent long 17 minutes learning about these characters in great detail. Much like the top-down knowledge of general film viewing, the use of a generic code, or a script, presupposes previous knowledge of the code or the script, thus assuming a reading and viewing practice that is already available cognitively. Literary critic Roland Barthes talks about the process of reading literature as always employing these kinds of top-down cognitive mechanisms:

---

[18] Schank and Abelson, op. cit., 55.

[19] M. French, Westerns: Aspects of a movie genre, Viking Press, 1973, p.23.

> The logic to which the narrative refers is nothing other than a logic of the *already read*: the stereotype (proceeding from a culture many centuries old) is the veritable ground of the narrative world, built altogether on the traces which experience (much more bookish than practical) has left in the reader's memory and which constitutes it.[20]

The genre piece then uses top-down expectations and plays with the perceiver's assumptions, but is also very much dependent upon the perceiver's previous knowledge (or what reception theorist Hans Robert Jauss calls the "horizon of expectation"[21]). The genre enables the filmmaker and the viewer to take short cuts, since much information is redundant, or already known. Schank and Abelson's model also accounts for those kinds of short cuts, or what they call causal chaining :

---

[20] Roland Barthes, The Semiotic Challenge, Hill & Wang, New York, 1988, 144.

[21] Jauss, Hans Robert, Toward an Aesthetic of Reception, Minneapolis: University of Minnesota Press, 1982.

> People, in speaking and writing, consistently leave out information that they feel can easily be inferred by the listener or reader.... If we hear that X caused Y, we must ask if X could cause Y directly and if it cannot we must figure out the intermediate events. This is the principle of causal chaining.[22]

While this model is very good in explaining the speed of processing of specific scenarios, it may be too local, and lean too heavily on top-down operations. After all, a first time viewer of a Western who happens to see *Once Upon a Time in the West*, is still more likely than not to understand the plot of the film. She may be surprised to see Bronson able to kill all three armed men, but as the next few encounters with him unfold, and because of his self respect, the music, his refusal to bend down to Cheyenne, and his overall demeanor, she will eventually learn that Bronson (nameless throughout the film) is indeed the good guy. How does this understanding come about if a script was unavailable? Is it purely by general narrative conventions, or are there any other clues, such as general human behavior, knowledge about the world, etc.? Is it possible at all that the film could be interpreted without any of these scripts or frames of reference at all? Intuitively, the answer to this last question is "no" in the case of the arts, and most of the cognitive literature support this claim. But even if we need certain kinds of knowledge prior to being able to perform any interpretation, we cannot assume that we know all, otherwise we will be always reading the already read, seeing the already seen, and we would lose interest in artistic texts. The script idea assumes rigidity and does not explain how a script can be modified (although Schank and Abelson enable some minor aspects of it to be modified). But a genre film always exists in dialogic relations with other films of that genre, often diverging somewhat from the very basic structure. While the structure may be repetitive and the characters redundant, some aspect of the film needs to be unique in order for the audiences to be interested in the film.

---

[22] Schank & Abelson, op. cit., 45.

How is this uniqueness manifested, and how does it affect our interpretive practices? In the case of *Once Upon a Time in the West* there is a unique deliberation in the first few important minutes of the film on the three characters who are negligible for the plot. This deliberation works against the generic conventions, by which we are usually introduced to the main characters of the film and to the main dramatic conflict, within the first five minutes of the film. Here we are introduced to one of the main characters after 12 minutes, and to the second, Frank (played by Henry Fonda) only half an hour later. An hour into the film we are still not sure why Bronson came to take revenge on Frank. As stated before, the perceivers are aware of this particular deviation only if they are aware of the convention.[23] But at the same time, to be aware of the deviation, one needs to be perceiving twelve minutes of bottom-up information which is slow, thin in narrative detail, and is completely irrelevant to the story that is about to be told. In other words, while the conventions of the genre may be a top-down, script like structure stored in our memory, the actual viewing of a film, and the enjoyment of it, is dependent on external (to our memory) and present tense perceptions. Schank and Abelson's model (as well as Minsky's) can only explain the bottom-up perception after it has been processed by a script. The problem then with the 'script' or 'frame' idea is that not only does it assume that we are always reading the already read, but it also leads to an infinite regress of experience, which can only be processed by way of a previous experience. But if the uniqueness of Leone's film can only be recognized during a comparison between the script and the actual perception, it seems that the tension between the top-down and the bottom-up needs to be explored further before determining a hierarchy by which every perceptual experience is processed through a script. One possible way out of the hierarchical cognitive architecture of the script theory is Jerry Fodor's theory of "the module."

---

[23] It is important to note that *Once Upon a Time in the West* does not only conform to the Western genre, but is also a part of a larger canon of artistic texts, the of the classical realist epic, in which events unfold much more slowly than in traditional Hollywood cinema. The tension between the two genres requires a doubly competent viewer, or one with at least two sets of schemata or scripts.

Jerry Fodor looks more closely at the relationship between perception and cognition at the time of the perceptual activity. Fodor notices that in some cases, despite of a cognitive belief, our perception leads us to a conclusion different than the belief. One of his favorite examples is the Müller-Lyer illusion in which two parallel lines are flanked by arrows pointing inward in one case, and outward in the other. Although the lines are objectively of equal length, the one with outward pointing arrows appears

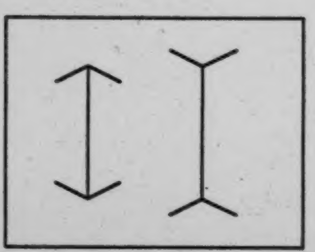


Illustration #1: Müller-Lyer illusion

longer. Even when one knows that the two lines are of equal length, they continue to look as if they are of different lengths. Fodor writes:

> Although perception is smart like cognition in that it is typically inferential, it is nevertheless dumb like reflexes in that it is typically encapsulated. . . The apparent difference in length of the Müller-Lyer figures, for example, doesn't disappear when one learns that the arrows are in fact the same size. It seems to follow that at least some perceptual processes are insensitive to at least some of one's beliefs.[24]

Because of perception's encapsulated nature, Fodor claims that it is distinct from cognition. In order for perception to be cognitively processed[25] it needs to go through (or be screened by) an appropriate *module*, which is defined as:

> . . .an informationally encapsulated computational system - an inference making mechanism whose access to background information is constrained by general features of cognitive architecture, hence relatively rigidly, and relatively permanently constrained.[26]

Fodor then identifies three different mechanisms that are active during the perceptive process: 1. transducers, 2. modules, and 3. central systems. The function of transducers is to receive energy impinging at the organism's surface and translate it into a representational form accessible by other psychological systems. The function of the central system is that of inference and belief fixation, and the function of modules is to mediate between the two. While this mediation may operate in either direction, Fodor discusses almost exclusively modules that take transduced representations and form hypotheses about their distal sources, hypotheses which then become available for use by central systems, which, in their turn, form beliefs and judgements, and draw inferences based on these hypotheses. Fodor says that the function of the modules is "to present the world to thought."[27] Perception then is modular, while higher cognition is unencapsulated, and has access to all modules and to the rest of the knowledge stored in the brain.

---

[24] Jerry Fodor, "Precis of the Modularity of the Mind," in Behavioral and Brain Sciences, 8:1-5, 1985, p. 2.

[25] Ibid. p. 3.

[26] Ibid. p. 3.

Unlike the script, a module refers to much larger units such as the syntactic or the semantic module, and is therefore less specific and more inclusive. This theory has several other advantages over script and frame theory: it assumes that at least at some basic level, perception operates independently of (sometimes even in opposition to) higher cognitive mechanisms. Unlike the subordinated nature of perception to cognition in the frame and script theories, Fodor points to a logic for a separation between the two systems. As a survival mechanism, perception needs to be specific, constrained and very sensitive to actual information from the world. If it relies too heavily on cognition, it may assume, or impose, certain beliefs on the situation, thus missing an accurate reading of the world. Interestingly enough however, Fodor too arrives at the conclusion that newly perceived information is processed in terms of a well-defined, contextually-based units of previously gathered information. Perception has to first be classified, or categorized so that the appropriate module can be picked out for operation, and only then, under the constraints of that module, can the transduced information be cognitively processed and affect belief by higher cognitive facilities.

---

[27] Jerry Fodor, The Modularity of Mind, Cambridge, MA: MIT Press, 1983, p.40.

One apparent problem with Fodor's theory, especially in its application to film perception, is that a module assumes compartmentalization in the perceptual activity, and, in its communication with high-order operations, and top-down cognition. Syntax, semantics, and vision are all processed through separate modules that have no access to one another. In *Once Upon a Time in the West,* for instance, there is a moment in which Bronson asks the three men why there is no horse for him. A module theory (*à la* Fodor) assumes that the sarcastic response "I guess we're short of one" will be processed independently, and without any low-level connection to the processing of the image and sound of Bronson playing the harmonica, and the shots of the three men already holding their hands on their guns. Such a theory would assert then that each of these linguistic, musical and visual events are interpreted through separate modules before they are integrated cognitively. But intuitively, such a separation will at best slow comprehension, and at worst hamper it (since body language, tone, and context will not be processed as a unified unit of signification). Such a limitation on the processing of multi-sensory information indicates that we need to look further to find an adequate model of film cognition and comprehension. And indeed, one area in which a worry about module theory is reflected in, is the debate between generative grammar theory (such as Chomsky's) and the cognitive grammar proponents (such as Lakoff and Langacker), which I shall turn to very soon.

Whether module, script, or frame, all these concepts assume, like Dretske's notion of meaningful perception, an a priori knowledge, or belief, about the object being experienced. It assumes that while bottom-up sensory information is being processed, it acquires meaning only when it is compared to top-down structures, which are already in place. It is relatively easy to show the frame, or script theory at work when discussing language. In order to understand language, we need to decode it, to compare it to a lexicon, or a dictionary stored in our brain. If we hear a word that does not exist in that lexicon, we would not know what it means, until it is explained or shown to us, or at least inferred from the context. But does this hierarchical processing schema work for images as well? We process a vast amount of visual information at once, in what seems to be parallel and simultaneous, more than hierarchical fashion. And more importantly, we rarely see an object twice from exactly the same position, distance, light condition, etc. How do we account for this when we try to compare it to a stored entry or category? Can visual categories work at all when each visual instance is unique in its individuality? Indeed, in the next sections I will show that (1) language research shows that high-order cognitive operations are always active in the perception of language, and (2) that the literature on visual perception presents some serious challenges to these hierarchical brain architectonics.

Language perception

Human language has developed as an auditory skill for communication long before it had a written form. In fact, anthropologists think that man has existed for at least one million years, and was able to talk through a great part of it. But the first evidence of written language appears only at 4000 B.C. with the writings of the Sumerians.[28] The nervous system specializing in language has developed as an auditory system, centered in the left hemisphere of the brain. Visually conveyed language is so new to human evolution that research suggests that there is no evidence of biological specialization for reading. The adjustments needed in order to accommodate visual language are intellectual, rather than evolutionary.[29]

Language is composed of different elements: the sound system, or language's phonology; the rules of word formation, or morphology; the rules of sentence formation, or syntax; and finally, the system of meanings, or semantics. Although each language has its vocabulary (a lexicon, or dictionary of words), all human languages have the same basic structural elements and in that structural respect they are all considered to be as complex. Human languages, considered as systems, are equally complex, no matter what the level of civilization of its speakers; Australian aboriginal languages, Amerinidian languages, and modern German all exhibit the same structural elements.[30]

---

[28] Philip Lieberman "The Originas and Evolution of Language" and Brian V. Street & Niko Besnier, "Aspects of Literacy," in Tim Inglod (ed.) Companion Encyclopedia of Anthropology: Humanity, Culture, and Social Life, New York and London: Routledge, 1994. Pp. 108-132, and p. 528.

[29] Thomas H. Carr "Perceiving Visual Language" in Handbook for Perception and Human Performance II, ed. by Keneth Boff, Lloyd Kaufman, and James P. Thomas, NY: John Wiley and Sons, 1986, 29:1.

Language processing functions linearly and analytically. Thomas Carr mentions that

> Language places a premium on dealing effectively with stimuli that are ordered in time. . . the information processing apparatus that subserves auditory language is highly adapted to working in a stimulus medium that consists of sounds ordered in time. Particular temporal relationship among particular sounds map onto more abstract phonetic, phonological, semantic, and referential concepts.[31]

Carr lumps together structural as well as semantic processing, as if they occur on a continuum. But the relationship between words and meaning is not nearly this simple.

Since the days of Aristotle and Plato philosophers and linguists have been discussing the nature of the relationship between the structural aspects of language (phonology, morphology, and syntax) and its semantics or meaning. Words stand as symbolic signs for concrete objects in the world, as well as for abstract concepts. There is relative agreement that the relationship between words and their reference objects is arbitrary and abstract. There is, however, much current debate about the nature of syntactic performance. I shall briefly describe these two areas of concern.

---

[30] Pinker, Steven, The Language Instinct: How The Mind Creates Language, New York: William Morrow, 1994.

[31] Thomas Carr, op. cit. pp. 29:3-4.

Up to the 19th century, language was considered to be a tool to express thought. In fact, Descartes' model of thinking assumed an immediate perception of ideas within the mind or soul. But linguistic structuralism, particularly the work of Ferdinand de Saussure, and pragmatist philosopher Charles Peirce challenged this notion, and changed the face of the debate. Both claimed that it is impossible to separate language from thought, and that language is not a tool to represent thought. Rather, Peirce and Saussure suggested that we *think in* linguistic signs, and they shape our thought process. Saussure said

> . . . it is impossible to isolate sound from thought, or thought from sound. (. . .) A linguistic system is a series of phonetic differences matched with a series of conceptual differences.[32]

And Peirce went on to say that "Thought which cannot be cognized does not exist. All thought must be in signs."[33] At the word level, both Peirce and Saussure point to a concrete relationship between the linguistic sign and what it stands for (not the actual object, but the idea of that object). This concrete relationship is symbolic, that is, abstract and arbitrary; there is no reason why d-o-g represents an animal that barks and not one that flies. The word "dog" is connected by convention and not by necessity to the animal that barks. And in order to semantically understand the lexicon of a language, we need to learn it, store it in memory, and match perceptual entries to the stored memory. In other words, an arbitrary and conventional sign system requires the user to encode and decode messages, clearly a high-order cognitive activity that involves the use of higher order mechanisms such as memory and categorization. It would seem fair to say that at the level of the lexicon, perception is always bound to be processed only via higher means of cognition.

---

[32] Ferdinand de Saussure, A Course in General Linguistics, trans. Wade Baskin, New York, NY: McGraw Hill, 1966, p. 166.

[33] Charles Peirce, "Questions Concerning Certain Faculties Claimed for Man," in Collected Papers, ed. Charles Hartshorne and Paul Weiss, Cambridge, MA: Harvard University Press, 1931.

But once we go beyond the word level, the nature of the relationship between a lexicon, a syntax and semantics is open for much debate. Noam Chomsky, who invented and was the central figure in the development of transformational generative grammar, claims that

> Among the formal structures are those of syntax, namely deep and surface structures; and also phonetic and semantic representations, which we take to be certain formal objects related to syntactic structures by certain well-defined operations.[34]

The grammar of a language, Chomsky claims, contains rules that relate syntactic objects (which are based on phonetic form and logical form[35]) to representations of meaning. The user of a language has "internalized these rules and makes use of them when he understands or produces."[36] In universal grammar (as well as in objectivist cognition) the assumption is that there is a separation between symbols and what they mean. Thought, under this view is seen as

---

[34] Noam Chomsky, Language and Mind, New York: Harcourt Brace Jovanovich, Inc. 1972, p.111.

[35] V.J. Cook Chomsky's universal Grammar: An Introduction, Oxford: UK, Basil Blackwell, 1988, p. 29.

[36] Noam Chomsky, op. cit. p.106.

> (. . .) the algorithmic manipulation of arbitrary symbols. The problem for such a view is how the symbols used in thought are to be made meaningful. The objectivist answer is that the symbols are meaningful by virtue of their association with things in the external world.[37]

Chomsky believes that all languages are similar in this structural respect (i.e., that there are rules that relate syntactic objects to representations of meaning), and he therefore postulates the theory of universal grammar. Furthermore, when looking at the fast language acquisition skill of children, Chomsky concludes that universal grammar is internal, a built-in capacity of our brain, which needs only be activated and used.[38] Under this theory, linguistic competence, or the knowledge of language is very different from how language is being used, or performance. Our mind is assumed to have a certain "black box," a syntactic module (to use Fodor's terminology), which processes low-order perceptual information, but is not affected, or changed by it.

While Chomsky is happy to "proceed with the study of 'knowledge of language' – what is often called 'linguistic competence' – in abstraction from the problems of how language is used,"[39] cognitive linguist Ronald Langacker argues that syntax is not as objective and pre-determined as Chomsky would like to portray it. Sentences like

---

[37] George Lakoff, "Cognitive Semantics," in Meaning and Mental Representations, eds. Umberto Eco, Marco Santambrogi and Patrizia Violi, Bloomington and Indianapolis: Indiana University Press, 1988, p. 125.

[38] Noam Chomsky, op. cit. pp. 112-113.

[39] Ibid. P.111. See also the distinction between grammatical competence and pragmatic competence, in Noam Chomsky, Rules and Representations, NY, NY: Columbia University Press, 1980, P. 59.

"the glass is half full" and "the glass is half-empty" may convey the same semantic information, but will create a different feeling or attitude, sometimes even a whole different context, and therefore different semantics.[40] Similarly, Langacker considers the semantic implications of the following sentences:

1. All cats are playful
2. Any cat is playful
3. Every cat is playful
4. Each cat is playful.

> These sentences share the conceptual content of a property (playfulness) being attributed to all members of a class (the set of cats). They nevertheless employ distinct images with respect to how one "reaches" or "mentally accesses" the class members for this purpose.[41]

Langacker draws attention here to the notion that semantics is not reducible to the connection between the signifier and the signified. All (as designating a group), and "every" (as designating all the members of a group), in sentences (1) and (2) above, create a very different mental image with regards to the group, and the relationship of its playful members to the group.

The issue of pragmatics of language use not only affects semantics, but may even affect syntax. Mark Johnson suggests that we conceptualize our experiences through the use of image-based schematas, such as containers, which determine the "in" and "out" relations of objects in the world. The most basic container is our own body, but the concept is extended to metaphorical uses such as "work-out" etc. Unlike semioticians like Peirce or Saussure, Johnson claims that we process our experiences prior to, and independently of, any concepts.[42] Bodily experiences are one of the fundamental ways of understanding, and postulating spatial relationships, and they affect our syntactics and semantics. Indeed, cognitive linguistics takes the experiential aspects of our life (including physical sensory and motor, emotional, and social) to be significant factors in the production and understanding of linguistic utterances. George Lakoff states that the central claim of experientialist cognition is as follows:

---

[40] Ronald W. Langacker, "An Overview of Cognitive Grammar," p.7.

[41] Ibid. p.8.

> Meaningful conceptual structures arise from two sources:
> (1) from the structured nature of bodily and social experience, and
> (2) from our innate capacity to imaginatively project from certain well structured aspects of bodily and interactional experience to abstract conceptual structures.[43]

While objectivist cognition sees the syntax of a language as arbitrarily determined independently of meaning, experientialist cognition sees the syntax of a language as providing grammatical categories that are semantically motivated. The relationship between form and meaning is much more complex in experientialist cognition; it is not seen as abstract or arbitrary, but as intertwined and dependent on experiential factors. Most importantly, language, in the cognitive linguistics view, does not function independently of visual, and other sensory practices.

---

[42] Mark Johnson, The Body in the Mind: The Bodily Basis of Reason and Imagination, Chicago: University of Chicago Press, 1987.

[43] George Lakoff, op. cit. p.121.

Using the research of Johnson, Rosch (to be discussed below), and others, Lakoff argues that kinesthetic, bodily and image-based schemata are all integral elements of any linguistic activity. Lakoff concludes:

> Cognitively real representations of meaning must make use of image- schemas. Image schemas are not finitary arbitrary meaningless symbols whose internal structure is irrelevant. Image schemas are nonfinitary (that is continuous), nonarbitrary, meaningful (via perceptual-motor experience), with a semantically-relevant internal structure.[44]

The notion that natural language processing uses at least some mental imagery in its semantic and cognitive processing leads us nicely to the discussion of visual perception. But before doing so I would like to point out that while cognitive linguistics has drawn attention to the fact that grammar does not function as an independent module, it, too, assumes a certain hierarchical and propositional processing. I shall show this when discussing Eleanor Rosch's formulation of basic level concepts, or prototypes.

It is also worth mentioning that when we use language, humans are capable of both receiving and producing utterances. And while the perception of language may not have much to do with the construction of sentences, there is something to be said about the greater understanding and competence in the use of language, that is a result of active participation in it. While we may all produce mental imagery, only very few of us produce public images for communicative purposes. The production of language is fast, easy and requires only the technology of our vocal cords. The production of images, on the other hand, is lengthy, requires skill and certain devices, which in the case of photography and film, are complex, technology based, and consume time. It is, naturally, very difficult to create an immediate visual dialogue, and only few members of society have a conscious understanding of what it takes to produce images. But we are all skilled in seeing images, and the next section will explain what that seeing involves cognitively.

---

[44] Ibid. p.149.

Visual perception

Seeing involves a neural response to light reflecting from the environment onto the retina in the back of our eyes. There are two kinds of light receptor cells in the retina, rods and cones. The rods specialize in seeing under light conditions of low intensity, such as dusk or night. Cones specialize in producing full spectrum of color under well-illuminated conditions. Our eyes can see in a range of about 180° but we see in clear focus only in the fovea, an area in the center of our retina that is about 2° wide. The fovea is mostly packed with cones, while rods dominate the parafovea, the near fovea and the rest of the retina. Robert Solso claims that

> Because sharp vision is restricted to a narrow band of available stimuli, we view objects, such as paintings, with eyes that are constantly refocusing on different regions. A consequence of this eye movement is that we do not see a painting all at once, as is commonly thought, but by forming an impression based on a large number of individual details falling within foveal vision.[45]

These constant eye movements and the re-focusing they involve are called saccades. Under this view we only see an image in its entirety once we reconstruct it as a mental image in our brain, which is, in a sense, a post-viewing product. Even the surprised close-up of the station manager, or the face of the gunman trying to blow off the fly in *Once Upon a Time in the West*, require time to be viewed as a series of saccadic eye movements, and to be reconstructed in our brain. Other researchers have shown that visual processing is not only executed in the foveal areas, but also in the parafoveal (and the peripheral areas of the retina), although the processing is generally not as complete.[46] In research on reading, Rayner claims that parafoveal and peripheral preview allow partial processing of upcoming text in advance of fixating that text and that this partial processing paves the way for more detailed foveal processing that ultimately results in word recognition. This pre-processing is generally considered to be not phonological (does not include uttering sounds) nor semantic (does not include processing at higher level of cognition), but is mainly visual (based on discrimination of features, etc). There is a clear connection here between low-level pre-processing of visual perception and high-order cognitive operations of text comprehension.[47] Such a view hints at more holistic attitudes toward image processing, to which I shall return later. This specific debate on the nature of images as a result of the

---

[45] Robert L. Solso Cognition and the Visual Arts, Cambridge, MA: MIT Press, 1994, pp. 23-24.

[46] R.N. Haber, "Control of Eye Movements in reading" and Hochberg J. "Toward a Speech-Plan Eye-Movement Model of Reading," in R.A. Monty & J.R. Senders (eds.) Eye Movements and Psychological Processes, Hilsdale, N.J.: Erlbaum, 1976.

[47] K. Rayner, "Foveal and Parafoveal Cues in reading" in J. Requin (ed.) Attention and Performance VII, Hilside, NJ.: Erlabaum, 1978, pp. 29-34.

capacity of the fovea and parafovea in the retina, is echoed in a much more philosophical discussion on the nature of our perception, the debate between direct realism and representative realism.[48]

From the retina, neurons simultaneously fire along two separate tracks to the lower back part of the brain, where the visual cortex is located, and visual processing is performed. The first, a ventral stream, is important for object recognition while the second, the dorsal stream, is specialized in determining spatial relations between objects.[49] Much of the research done focuses on object recognition and naming, thus on the ventral stream. But in our daily experiences of our environment, and in the perception of moving images, the relations amongst these objects, or the function of the dorsal stream, are as important as object recognition. Even more important for the case of film is the fact that the ventral stream may sometimes not be sufficient, since the object to be recognized has never been seen before. The *Alien* example mentioned earlier is such a case: comprehension cannot be based on the ventral stream, but relies on the dorsal stream (where is the alien coming from?). As a result of the inability to rely on the ventral stream, the perceiver also needs to rely on higher cognitive processing of the emotions and responses of the other characters in the film. While interpretation at large may require recognition of objects through the ventral stream, the suspense in the *Alien* relies precisely on the inability to make conclusive judgements about object recognition. The Alien is seen looking similar to a lobster, a snake, a sea cucumber, a robot, plain teeth, and eventually all this mass can be changed so as to be airborne through the ventilation system. The inability to reach a conclusive decision about the alien's shape leads to reliance on other available information, such as the behavior of the characters involved, generic conventions and expectations, etc. The interplay between the audience's insecurity about the Alien's form, and the interpretation of the overall narrative consequently relies on the interaction between low-order visual processing (both the ventral and the dorsal streams) and high-order problem solving mechanisms. I will therefore now turn to briefly describe the research on the ventral and the dorsal streams.

---

[48] Both doctrines share the notion that there is a real physical world, which is independent of our perception of it. But as Dretske puts it:

> Direct realism holds that. . . under normal conditions observers are, in a direct and unmediated way, perceptually aware of the objects and facts that constitute this world. (. . .) According to representative realism, our perception of physical objects is indirect, mediated by a more direct apprehension of internal representations . . . of external physical objects.

For representative realism then, knowledge about the objective physical world is available to us only through subjective mental facts and internal representations. Solso, with his "after the fact" image is a representationalist, while Gibson (ecological approach) is a direct realist.

In the case of film, such a debate becomes even more complicated. As mentioned earlier, the viewer is aware that the pro-filmic event happened in the past, while shooting the film elsewhere, and that it seems "real" only through a manipulation of optical illusion (both photographic and editing). And indeed, much of the cognitive film literature is devoted to an exploration of the nature of this illusion. But whatever kind of illusion film is, the film images and sounds are either experienced directly at the time of viewing, as an unmediated perception of a screen image in a theater, or the filmic construction is perceived representationally, via some mediated mental capacity. I shall not attempt here to resolve the debate between direct realism and representational realism, since it is only indirectly relevant to the question of perception and cognition of actual filmic information. But I wanted to mention this debate here, since it is the main locus of work for so many cognitive film theorist.

[49] Nakayama, Ken & Zijiang J. He, and Shinsuke Shimojo, "Visual Surface Representation: A Critical Link Between Lower Level and Higher Level Vision," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995, p.3.

The Ventral stream

Most research done on object recognition is done in a laboratory situation where subjects are asked to perform an array of tasks such as identifying whether a target object is in a scanned group, and matching shapes, or letters and digits to scanned groups. These experiments in extremely controlled circumstances have fascinating results as far as object recognition goes.

At the very basic level, Neisser notes that

> Visual search involves parallel processing. During each fixation, subjects must (1) decide whether the item they have fixated is a target (2) select the next item in the periphery to fixate, and (3) organize a saccadic eye movement to bring the fovea to bear on the next item.[50]

When looking at how we come up with a decision about whether an item is a target or a non-target, Neisser came up with a hierarchical processing schema:

> Stimuli are first tested for lower order physical properties in parallel, and non-targets may be categorized at these lower-order levels. That is, if the stimulus fails these lower order tests, it is classified as a non- target without further processing. If the stimulus passes those lower order tests, higher order tests are conducted on more abstract features of the stimulus.[51]

Neisser's research then shows that two visual search processes are active at any search: preattentive, and focal attentive. The preattentive search is a crude scan of the whole visual field for easily discriminable physical features. Focal attentive search, on the other hand, is a slower serial mechanism operating on one object at a time. Non-targets get a preattentive treatment, while targets get the focal processing. But this division only works for distinct non-targets. If the non-target is relatively similar to the target it gets full focal processing. Chase and Cavanagh posed a challenge to the low-level, shape based, binary-processing thesis of Neisser's. By showing that color, size and shape all affect the identification of targets or non-targets, they show that pre attentive and focal attentive processes can happen for different aspects of the object, and not always in accordance with one another.[52]

---

[50] U. Neisser Cognitive Psychology. New York: Appleton- Century-Crofts, 1967. p.282.

[51] Ibid. p.288.

The decision whether an item is a target or a non-target is made by the use of template theory which "postulates that pattern recognition involves matching sensory input against specific, labeled template like representations stored in memory."[53] There are two major problems with the template theory, though. The first is that for recognition to occur there must be a match, which means an already perceived memory of that object. Secondly, and more importantly, Neisser's theory is based on the assumption that a vast number of memories need to be stored as templates so as to tolerate size changes, and different perspectives on three-dimensional objects. In other words, the templates are fairly specific. Such a theory would claim that recognizing the gun resting on the leg and pointing down in the tilt-up shot at the beginning of *Once Upon a Time in the West*, and recognizing the gun used to trap the fly a dozen shots later, require two different templates. But prototypical templates can relieve some of the pressure on memory requirements.

---

[52] William G. Chase, "Visual Information Processing," in Handbook for Perception and Human Performance II, ed. by Keneth Boff, Lloyd Kaufman, and James P. Thomas, NY: John Wiley and Sons, 1986,p. 28-26.

[53] U. Neisser, op. cit., p.156.

Posner and Keele were investigating whether our recognition of objects is aided by a processing that takes into account prototypical shapes.[54] In an initial experiment they showed subjects figures which were distorted out of a prototype of a triangle (so that they did not resemble triangles), and none of the subjects named any of the distorted shapes - "a triangle." In a second experiment, the same subjects were shown some of the old shapes, plus the prototype (a triangle), and some new shapes. The subjects immediately recognized the prototype, named it a triangle, and then matched and compared all other shapes to it, as good or bad examples of the prototype. The more similar a new pattern was to the prototype, the quicker and more accurately it was classified. Posner and Keele concluded that the more experience we have with real-life encounters of a category, the less we remember individual instances, and the more we rely on the general (prototypical) properties of the category. They define the prototype as "an averaged, modal (no extreme examples), or the most typical pattern out of a set of instances."[55] The prototype approach is very efficient as far as memory goes, and therefore addresses the first difficulty with Neisser's template theory, that of storing individual instances in memory. But the second problem with Neisser's theory, that of the specificity of each image, is not resolved by this model either. Distance, color, lighting conditions, angle of viewing, and occlusion all affect how we perceive an object. Indeed, there is little in common (visually speaking) between the gun tucked in the belt, pointing down and slightly occluded by the duster, and the sideways view of the gun held in front of the gunman's face with the fly in it. It seems hard to make a case that there could be a single visual prototype that will account for both instances of the category "gun." But it makes sense that there would be a higher-level category "gun" that could be imposed on both instances. This abstract general concept will, of course, require social, historical and functional knowledge that will help compensate for the visual differences.

---

[54] M.I. Posner & S.W. Keele, "On The Genesis of Abstract Ideas," Journal of Experimental Psychology, 1968,77, 353-363.

[55] Ibid. p. 356.

The work on visual prototypes (especially these higher-level categories) is closely related to the work on linguistic prototypes conducted by Eleanor Rosch and others.[56] Rosch noticed that some instances of a category are considered more typical or better examples of that category. For instance, subjects were more easily identifying a robin as a bird than a chicken, and their identification time was 10% faster when a typical instance of a category was presented. Rosch used the (Wittgenstinian) concept of family resemblance within a category: typical or prototypical members share more attributes with other members of the category, but at the same time they provide maximal discrimination from closely related, contrasting categories.[57] Categories of real world objects are hierarchically structured, and there is generally a most basic level of abstraction, or the prototype. The basic level bird, for instance, is placed in between the superordinate (animal) and the subordinate (robin) levels of abstraction and specificity. Rosch et al. claim that

---

[56] E. Rosch, "On the internal structure of perceptual and semantic categories," In T.E. Moore (ed.) Cognitive Development and the Acquisition of Language, New York, NY: Academic Press, 1973.

[57] E. Rosch and C.B. Mervis, "Family resemblances: Studies in the internal structure of categories," Cognitive Psychology, 1975, 7, 573-650.

> Basic categories are those that carry the most information, possess the highest category cue validity, and are, thus, the most differentiated from one another.[58]

In accordance with Gleitman & Gleitman's research (mentioned above) on language acquisition of young children, Rosch shows that basic level concepts (for instance "chair") develop at an earlier age (3-5), than superordinate ("furniture"), and subordinate ("rocker") ones (5-8). Important for the discussion on vision, Rosch notes that basic levels are the highest levels of abstraction for which people can generate images to facilitate perceptual and motor processes. So if one hears "animal" (superordinate) one is at a loss in coming up with a mental image, but if one hears "bird" one comes up with an image immediately and efficiently. But it is, of course, a very different issue for subjects to *produce* a mental image of an object after receiving a basic level verbal cue, than to *use* prototypes while perceiving visual information. In other words, Rosch moves back and forth between language and images without acknowledging the difference in perception, and possibly also in the production of mental imagery, versus that of a natural language. And while I support the cognitive linguistics' claim that the processing of both language and images is inter-dependent, it is important to deal with the peculiarities of the perception of each medium, and only then (and quite carefully) chart the relationship between them in other cognitive processes.

---

[58] E. Rosch, C. Mervis, W. Gray, D. Johnson, & P. Boyes- Braem, "Basic Objects in Natural Categories," Cognitive Psychology, 1976:8, p. 382.

Geons

While trying to apply the prototype theory to images, and solve the problem of how subjects maintain invariance of concept over changes in size, point of view, and other variables, Irving Biederman comes up with an interesting solution. Biederman noticed that "visual entities almost always invite a decomposition of their elements into simple parts."[59] The manner of the decomposition does not depend on subjects' familiarity, or recognition, of the object: Biederman's research shows that "nonsense shapes" were decomposed to similar subshapes by subjects. These subshapes are based on viewpoint invariant properties, such as parallel lines, curves, straight lines and symmetry. Biederman then

> proposed a theory of entry level object recognition that assumes that a given view of an object is represented as an arrangement of simple, viewpoint-invariant, volumatic primitives called *geons*. (. . .) The geons have two particularly desirable properties: they can be distinguished from each other from almost any viewpoint, and their identification is highly resistant to visual noise.[60]

Biederman identifies twenty-four geons, such as brick, cylinder, and cone, which can be set in different relations and aspect ratio to produce 10,497,600 possible two geon objects.[61]

---

[59] Irving Biederman, "Visual Object Recognition" in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995, p. 129.

[60] Ibid. p.139.

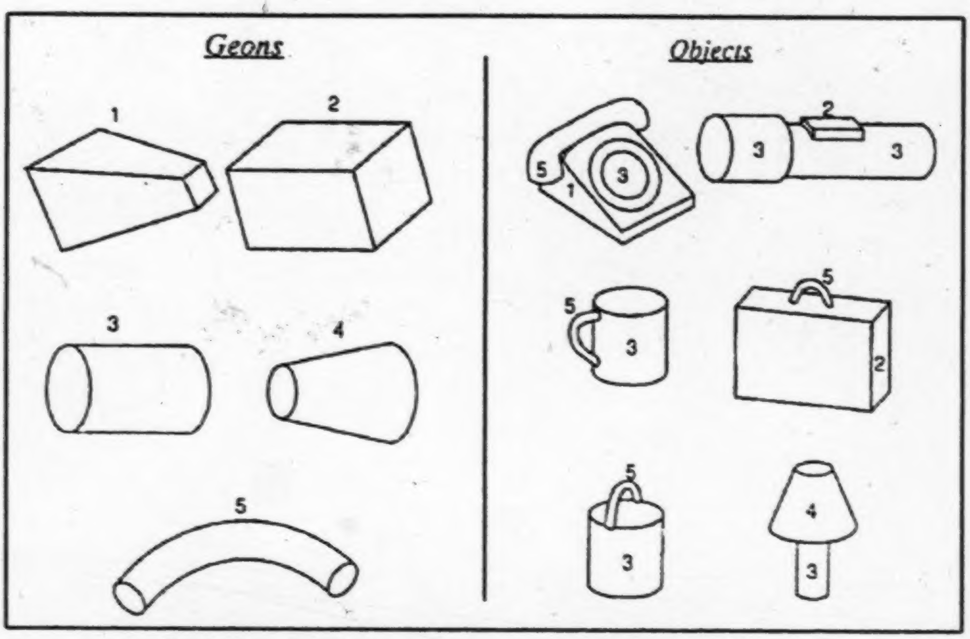


Illustration #2: Biederman's decomposition into geons.[62]

In a series of experiments Biederman and others examine the theory of decomposition into geons, and the results are overwhelmingly in support of the theory.[63] Biederman writes:

> The theory thus implies a *principle of geon recovery:* if an arrangement of two or three geons can be recovered from an image, objects can be quickly recognized even when they are occluded, rotated in depth, novel, extensively degraded, or lacking customary detail, color, and texture.[64]

---

[61] Ibid. p.143.

[62] Ibid. p. 140.

[63] One may also think of "how to draw" instruction books that simplify complex objects into a set of basic, geon-like shapes.

[64] Ibid. p.144.

It is interesting to note that the first gun seen in *Once Upon a Time in the West,* although partly occluded, reveals the two important geons: a lengthy pipe like cylinder, and the ring of the trigger. Were the trigger hidden under the coat, we would have been left with one geon, the cylinder, and would have had more difficulty recognizing the object at hand. The geon theory also explains why the audience has such a difficulty anticipating the presence of the Alien in *Alien*. Given that the filmmakers kept changing the size, shape, and material of the Alien, and given that often only a small part of it shows (in a close-up), or it is occluded, it is impossible to decompose the Alien to a series of basic geons.

Moreover, a theory of decomposition to such basic forms does help us retrieve and store a representation in memory, under varied visual conditions. Unlike Neisser's templates, or Posner and Keele's prototypes, Biederman's theory is a bottom-up theory. Images are decomposed into basic shapes (geons) first, and only then are recognized. Different experiments conducted by Biederman show that the priming in these experiments was largely visual (based on the shapes and natural decomposition into those shapes), rather than contextual or lexical (i.e., based on a pre-conceived concept).[65]

Biederman also points out that even in the absence of any context, the speed of object recognition is not much slowed down, data which argues against a central role for top-down mechanisms such as categorization. This research suggests that the initial perception of visual objects is done in a mainly bottom-up manner, and only when naming and classification are involved do higher order cognitive activities come into the picture. Biederman's research also enables us to come up with a model for visual memory storage. Instances are abstracted to geons which compose prototypical (Rosch-style) entries, but these prototypes are easily manipulated (the variant aspects of the geons, like size, rotation and relation to one another), so as to be matched to particular perceptual instances.

---

[65] Ibid. p.149.

Research on the ventral stream then, indicates a specific interaction between higher-order cognitive mechanisms and low-order processes during the time of perception of particular objects. While bottom-up information is processed, it is mostly classified and stored in categories. Whether using templates, prototypes, or decomposition into geons, the low-level visual perception operates by comparing sensory input to existing data in our memory. But the ventral stream research focuses on object identification, and during our daily activities we only sometimes search for a specific target in our visual environment. We much more often process the environment without a specific goal. In these cases our eyes are overwhelmed by a multiplicity of stimulus, and the speed with which we process those complex images indicates that it is unlikely that we decompose each and every object to determine what it is, before deriving a meaningful perception about the whole image. Research on the dorsal stream can therefore shed some light on how we process complex visual environments.

The Dorsal Stream

As mentioned earlier, the dorsal stream is responsible for determining spatial relations, mostly as a guiding tool for motor interactions. Here position, size, and orientation in space must be determined with great accuracy. In the following pages I will discuss some of the research done on the interpretation of complex images such as those we are faced with in our daily physical activities.

Developmental psychologist Jean Piaget explains that an adult-like representation of large-scale environments develops only at ages 11-12.[66] Up to the age of six or seven children are at an egocentric (or proportional) stage: they locate objects in the environment relative to their own bodies, and using their bodily experience they learn to evaluate proximity, separation, open and close shapes. In other words, objects may seem larger or smaller depending on the child's distance from them, and the child is unable to compute the distance/ movement factor into a stable representation of the space. At ages seven to nine children move into the fixed (or concrete operational) stage, where a fixed coordinate system -- in which objects and one's own body are oriented relative to fixed points and landmarks in the environment -- is established. Here the child orients itself based on stable objects, and is able to comprehend distances and her own movement relevant to these objects. At this stage more complex topological properties are perceived, such as "between," order, and enclosure. Around age ten or eleven children reach the coordinate (or formal operational) stage. At this stage an abstract coordinate system, such as the cardinal system, develops, and the topological properties of continuity and other abstract relations are established. Now the child is able to have a mental floor-plan of the environment, one that is independent of herself, and her placement in space. It is interesting to note that until the age of eleven we do not have an abstract propositional concept of our visual environment, but one that is changing and is relative to our body's and other objects' positions in space.

---

[66] J. Piaget & B. Inhelder, The Child's conception of Space, New York, NY: Norton, 1967.

Visual processing of large spaces requires several sub-processes. Treisman, for instance, looked at how subjects group together objects that share certain features, as a way to process the whole image faster and more efficiently.

> A perceptual group is a collection of spatially proximate objects which share a preattentive feature in common (shape, size, color, texture, orientation). The interesting property of a group is that focal attention can be brought to bear on the group, and the group as a whole can be processed in parallel with respect to the grouping feature. Spatial proximity is important because if two feature groups are intermixed, attention can be focused only on individual objects.[67]

Treisman's results show that if the target is unique, and different from the set, it can be searched by "feature search" which is much like Neisser's pre-attentive process, fast and parallel for a few groups. That is, when the target is unique, a crude search on whole groups can detect the presence or absence of the target. But if the target shares a feature with the set, there is a conjunctive search, which is focal, serial, and controlled. That is, a slow scan, searching each item in the set has to be performed in order to verify the presence or absence of the target.

---

[67] A. Treisman, "Perceptual grouping and attention in visual search for features and for objects," Journal of Experimental psychology: Human perception and Performance, 1982, 8, p. 198.

Treisman's experiments included placing triangles in the midst of rectangular groups, or groups of triangles only, where a few are positioned differently than the group. This is, of course, quite a low level processing of simple images. But I believe that the idea of grouping can be applied to much more complex visual levels. In *Once Upon A Time in the West* we see all three gunmen in the same frame on two different occasions. The first is in the fourth shot of the film, after we saw the scared look on the face of the station manager. The film then cuts to a shot (presumably from the manager's point of view) panning through the room showing all entrances to the station blocked by the three men. They are all dressed in dark with yellow dusters and black hats, and even though they are on the right, middle, and left side of the frame, they stand out from the brown wooded station walls, as a group. The scene progresses with mostly individual shots of each of the gunmen while they are waiting. The next time we see them together it is from the point of view of Charles Bronson: three darkly silhouetted figures on the station's platform, turning towards the sound of Bronson's harmonica. Leone then alternates through the dialogue in a series of close-ups, at which end the three gunmen are seen as a group one last time, just long enough to be shot by Bronson. While most of the twelve minutes were devoted to individual shots of the three men waiting, group processing is at work, not only visually, but also psychologically, as their menace comes from their identical behavior, similar clothing, and overall demeanor. Moreover, their importance in the context of the whole film lies solely in the fact that they were eliminated (as a group) with no great difficulty by Bronson. In other words, in retrospect, after the viewing is completed, the twelve minutes we spent studying those men individually were narratively misleading and useless, as we did not need to know them at all. All we needed to know narratively is that they were three, that they were rough, and that they were all killed by one fast-on-the-draw man. Their overall narrative importance then lies only in their shots as a group, which interestingly are provided from the point of view of other characters. The sequence, of course, has other significance, as it sets the mood for an epic, prepares us for the aesthetics of the film, and places it with relation to the genre as a whole. While it is thin in narrative detail, the scene is rich in atmosphere and other important elements to the telling of the film. Visual processing here (a focal group) echoes narrative processing (a thematic group), and allows for attention to be placed on the epic qualities of the film.

The discussion of grouping, which involves processing several objects of the same group at once, brings us to a related issue, that of holistic processing. Martha Farah noticed that

> Whereas most objects are only a bit harder to recognize upside down than right side up, inversion makes faces dramatically harder for normal adult subjects to recognize.[68]

By running a series of experiments with patients who suffer associative agnosia (an impairment of visual recognition), she managed to determine that some patients had object agnosia, but no impairment with faces, and some had the exact opposite syndrome. As a result she concluded that there are two specialized visual recognition systems, one for objects and one for faces, which operate independently of one another in a modular fashion.[69] Her experiments with visual agnostics led her to conclude that face recognition is a holistic process, which require no decomposition into parts, and therefore requires the ability to represent complex parts. Farah concludes that

---

[68] Martha J. Farah, "Dissociable Systems for Visual Recognition: A Cognitive Neuropsychology Approach," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995, p. 101.

> Face recognition and common object recognition depend on different systems that are anatomically separate, functionally independent, and differ according to the degree of part decomposition used in representing shape.[70]

This research then suggests that while watching a film we use two different visual processes of perception: one holistic to account for faces in the frame, and the other decompositional (in a similar style to Biederman's), which analyzes the space around the characters.

So far, I have been discussing the perception of one image, a frame or a shot of a film. We have just seen that there are at least two operations at work (holistic and decompositional) in the perception of each shot of the film. The balance between these processes will probably depend on the kind of the shot (a continuum from an extreme close-up of one object to a wide-angle shot of a large space) and its duration on the screen. *Once Upon a Time in the West* provides many lingering close-up shots, enough time to thoroughly examine the frame and complete a Biederman style decomposition. But *Alien* provides only a very few, short, and obscured shots of the Alien, so as to prevent the audience from carrying out the kind of thorough processing needed for comprehension. It will be interesting to find out whether the perception of images is strategized so as to first complete a crude, holistic scan of the image, and then, depending on the time available, process details in a careful focal attentive style. And to date, I have been unable to find data about experiments that determine such strategies for image processing. But whether holistic or focal attentive searches, the shots are not interpreted as static images, one at a time, nor as hermetic units. Film theory acknowledges that every image is processed in relation to the others in the dramatic unit, the scene. A scene, which is defined by a dramatic unity of place and time, usually involves editing, which shows us a multiplicity of visual perspectives in a variety of shots. Generally, the audience of a film tries to come up with a "mental floor plan" of the scene, so as to place characters in relation to one another.[71] Most traditional narratives will provide early on in a scene a wide-angle shot, one that establishes the space and the spatial relations of the characters in it. This wide shot (also called establishing shot) helps the audience expand the visual space when processing close-up shots and details of action. Editing, then, presents a different challenge as well: the understanding of the overall space of the narrative world. By matching the movement of characters over the edit, and by the use of filmic conventions that have developed in the past one hundred years, most of the editing in conventional cinema seems seamless to the audience. Yet, in order for this editing to make sense spatially, so that we can accurately predict where characters will come from, we need to reconcile the perspectival changes, expand and alter the spaces we have seen, and obtain an overall "floor plan" of each scene. A useful model for understanding how we come up with a concept of the space of a scene can be found in research on cognitive maps.

---

[69] Ibid. p.103-110. See also Mark H. Johnson and John Morton Biology and Cognitive Development: The Case of Face Recognition, Oxford UK & Cambridge, USA, 1991.

[70] Martha J. Farah, op. cit. p.118.

---

[71] In Narrative Comprehension of Film Edward Branigan provides thorough analyses of different scenes, which include shots, camera angles, and floor plan explicating what the audience sees, and how these visual structures restrict and determine narrative knowledge. See Edward Branigan, Narrative Comprehension and Film, New York and London: Routledge, 1992.

Large-scale environments cannot be viewed from a single vantage-point to provide all the information about elements in that environment. Some objects will, by necessity, be occluded by others, and therefore knowledge about these environments must be at least partially inferred rather than perceived. It is a common practice to apply the principle of closure to shapes that are partially occluded, and to assume that they continue behind the occluding object.[72] Given that defining surface continuity is crucial for object recognition, it seems that already in a very early stage of our perception we infer, rather than perceive, information about invisible (from that vantage-point) entities. This, of course, presents an epistemic problem of how we actually know about our environment. It is safe to assume that most of these inferences are made based on our experience with the real physical world, and therefore they take into account laws of gravity, the nature of our perspective, the work of light rays, etc. Similarly, in film, we are provided with bits and pieces of spaces (shots), clearly outlined by the borders of the frame. Yet, we never assume that the spaces end at the frame line. Rather, we infer a continuation of the space beyond the frame to connect the different fragmented shots that comprise the scene. The massive use of close-up shots in the opening twelve minutes of *Once Upon a Time in the West* prevents us from identifying those spatial relations, but as it turns out, when the important action begins we are provided with a long shot of all three gunmen, and then the camera cuts to an over the shoulder shot of the gunmen looking at Charles Bronson, so that we can have a good sense of the spatial relations between those characters. Interestingly, when the train starts pulling away, and before the gunmen realize Bronson is there, there is a shot of all three taken from above (a bird's view), which cannot be justified as anyone's point of view. This shot gives us an overview of the space, almost like a floor plan, or a map, so that when action starts we are well aware of the spatial relations.

---

[72] Nakayama, Ken & Zijiang J. He, and Shinsuke Shimojo, op. cit. P.2.

Already in 1948 Tolman had shown that rats develop a two-dimensional map of their environment.[73] Robotics research conducted by Winograd shows too that an internal representation of the environment, and a representation of the location of the robot with respect to this environment is essential before movement can be planned.

> The meanings of actions and statements about the environment must be reduced to descriptions of this internal representation and, further, proposed actions must be "simulated" in the internal representation to see if they are possible and to derive a plan of action to be executed on the external world.[74]

Winograd claimed that this internal representation is a model, which can be propositional and does not necessarily resemble the environment pictorially. But this discussion about abstracting spatial knowledge developed into a debate over whether visual images have the properties of an analog representation (such as a map), or whether they have the properties of propositional representation (such as verbal description). Levine found out that people who learn an environment by experience develop an internal map, which is more accurate as regards to routes (i.e., they remember all landmarks, turns, etc.). But as far as orientation goes (north/south, etc.,) this map is less accurate than the map of those who learn an actual map of the environment first. Map learners perform spatial tasks by visualizing a mental image of the map they have learned, whereas people who learned the area by direct experience perform these tasks by mentally simulating a walk through the area.[75] This data supports a claim that at least in some cases the representation is analog or depicitve.

---

[73] In William G. Chase, op. cit. p.28-39.

[74] T. Winograd, "Understanding Natural Language," Cognitive Psychology, 1972:3, p.125.

Levine has concluded that there are two properties of mental maps that are essential to an analog representation (as opposed to a propositional one): triangulation and rotation. Triangulation postulates that if the locations of A and B are known, and the relation of the locations B to C is known, the relations between locations A and C can be deduced. Spatial triangulation has an advantage over a propositional scheme, which would require a long chain of inferences to take place before these relations are figured out. Rotation postulates that analog images have a preferred or canonical orientation (up and down in particular), much like Farah's observations about holistic perception of faces being interrupted by rotation.[76] But Tversky found much evidence in support of propositional representations. She first noticed that people tend to abstract mental maps along a grid system. When asking subjects about the spatial relations between San Diego and Los Angeles, most of the answers were based on the north-south axis, even though Los Angeles is further west than north of San Diego. Subjects aligned the two cities in memory along north-south, east-west axis, for storage purposes, and Tversky determined that there is encoding or alignment error for storage purposes.[77] This rotation error can be explained as a gestalt process of aligning a map in memory with an already existing grid of reference. In other words, establishing a mental map, even a brand new one, is done in reference to a strong top-down schema of a grid system. Tversky also found that expert taxi drivers and novices alike produced the same spatial distortions when asked to provide a two-dimensional map of the cities they were driving in. But experts were much more likely to come up with an efficient route between two locations. They first found a connecting route between the location and the destination, and then found connecting routes to the major linking route. Locations then, are nested within neighborhoods, neighborhoods within larger regions, etc. This evidence shows that there is a propositional network -- which is structurally based on nesting -- at work when choosing a route.[78] Whether the initial representation is analog or propositional, both Levine and Tversky agree that long-term memory is categorized and

---

[75] M. Levine, "You-are-here maps: Psychological considerations," Environment and Behaviour, 1983:16, p.139-157.

---

[76] Ibid.

[77] B. Tversky "Distortions in memory for maps," Cognitive Psychology, 1981:13, pp. 407-433.

[78] Ibid.

summarized. We do not store directly all possible spatial relationships, and therefore mental images of spatial relations are by necessity impoverished and schematic. However, more recent research on mental imagery offers an interesting picture about the nature of stored images and maps, one that reconciles the analog and propositional views to an extent, and I shall return to this view shortly.

Before closing this section I would like to briefly discuss the issues of parallel processing and divided attention. Harold Pashler researched what happens when subjects are presented with several visual stimuli and attempt to perceive them all.[79] Pashler found that there is a capacity limit on parallel processing, as there is a short-term memory used for such processing, which cannot hold more than four or five items. Moreover,

> There is a postperceptual bottleneck: when a person retrieves a response to one stimulus (or engages in memory retrieval. . .), he or she cannot retrieve anything else at the same time.[80]

Fast channel changes on the TV remote control, for instance, produces a sense of understanding each image, but a difficulty in reporting back on what was seen. Pashler's concludes that while we can perceive and analyze more than one object at a time (within the same modality), we don't seem to be able to store these outputs in memory, unless they go through careful serial processing.

The implications of this research to understanding the perception and cognition of film, particularly of mainstream drama, is enormous. At the actual time of perceiving a scene we are mostly busy in determining spatial relations and anticipating future actions. Like Levine's subjects who first learned an environment, and then simulated a walk through it, we are progressively learning the environment, as if we are in it, behind an invisible window. Our initial perception may therefore be holistic and analog, and its extent will depend greatly on the exposure time to each shot. But in order for this material to be used later in the reconstruction of the narrative it needs to be stored in long-term memory, and for that purpose it is summarized, categorized, abstracted and probably transformed into serial propositional representations.[81] I shall now turn to the nature of stored visual information.

---

[79] Harold Pashler, "Attention and Visual Perception: Analyzing Divided Attention," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995, p. 74.

[80] Ibid. p.93.

Mental Imagery and Memory

In the research on memory there is a distinction between working memory (or short-term memory), and long term memory. Parallel processing of the kind Pashler talks about is possible in short-term memory, which is active while the image is being perceived. Similarly, research on working memory during the perception of language shows that knowledge of syntax is used to parse the surface structure information into major syntactic constituents: phrases, clauses, and sentences. Working memory can hold 7±2 items, which are processed into a group (that can then be processed as one unit in larger groups), and the content of the constituent is computed and added to the text representation being stored in long term memory.[82] Long-term memory, as discussed by Pashler, need condensed (that is stereotyped or schematic) and encoded (i.e., propositional) messages as its material.

---

[81] One of the painful lessons I learned as a filmmaker was that the end of each scene needs to include a closing shot (or part of one), a time in which no dramatic action takes place, but that gives time for the integration of the new information obtained, before moving on. When I failed to do that, the audience got frustrated, as if not having enough time to "get" everything that happens in front of their eyes.

While looking at the acquisition of cognitive skill, Anderson came to an interesting conclusion about long-term memory.[83] Anderson identified two components to semantic memory: 1. declarative knowledge, which is stored in a propositional network of facts, and 2. procedural knowledge, which is stored as a set of symbolic condition-action sequences called production rules. Learning involves acquiring facts, and converting them to production rules, and this proceduralization underlies the automation of skills, which means a reduction in memory loads. Speedup in complex cognitive skills is the result of two processes: 1. composition of production rules, which are sets of rules that always follow one another, and 2. increases in strength of production rules. The time it takes to execute a production rule is a direct function of its strength, which is a result of each correct application of the rule. The facts about the changing size of a moving car, for instance, are transformed to rules that enable us to cross the road safely. It is likely that an urban person will be able to decide faster than a rural person who lives away from traffic, whether it is safe to cross the road or not. Or put another way, it is likely that the production rule of crossing roads with traffic is stronger for the urban person.

---

[82] T. H. Carr, "Consciousness in Models of Human Information Processing: Primary Memory, Executive Control, and Input Regulation," in G. Underwood & R. Stevens (eds.) Aspects of Consciousness, London: Academic Press, 1979.

[83] J. R. Anderson, "Acquisition of Cognitive Skill" Psychological Review, 1982: 89, pp. 369-406.

Indeed, research on professional chess players reveals that the masters' general visual memory is not better than beginners', but their ability to remember contextually-based, stereotypical positions is much better.[84] When shown a mid-game position, 75% of the masters' reconstruction skill was based on stereotypical positions, and the rest on grouping of color or shape. While the beginner and the master both exhibit the same ability with regards to short-term memory, it seems that the master has a large number of prototypical positions stored in long term memory, which the beginner does not. Beginners tend to analyze every move, but with training subjects moved towards a perceptual approach, looking for familiar patterns with familiar best moves.[85] This data makes it clear that the success of master chess players arose from their ability to turn facts of the game into production rules, which usually determine a whole series of moves, thus explaining the ability of master chess players to play at once against dozens of people.

In the case of film, this research has implications on various levels. First, on a basic level, we do not process each new filmic space as novel, but we instead apply production rules from our life and film experience (not many of us have been to a Western saloon, yet we recognize it as if we have!) to the processing of familiar spaces. This increases the speed of processing, and enables us to concentrate on the important narrative aspects of the film. Secondly, on a generic level, much like the chess players, some generic conventions function like production rules. The production rule of the Western (and possibly of drama altogether), determines that we know that since Charles Bronson killed Henry Fonda's men, Fonda is bound to retaliate. There need not be any specific mention of that in the body of the film, as this information is already present for the competent Western viewer. On a narratorial level too, then, at least some of the structure of the plot is known and already processed, thus not requiring a large cognitive effort to process and comprehend. Artistic texts are free, of course, to break those rules, and some may even say that most of the interest in generic cinema arises from the tension between the generic code and its violation. An example can be seen in one of the last scenes of *The Silence of the Lambs,* in which director Jonathan Demme uses a conventional parallel editing technique (cutting between two places back and forth, therefore indicating that both scenes are happening simultaneously). In the climax of the scene we see the FBI agents ring the bell at the door of the serial killer.[86] The camera cuts inside to the distressed killer hearing the doorbell, and getting ready to answer the door, indicating that both actions have been united in space and time. But when he finally opens the door, we see Claris (Jody Foster) who supposedly is thousands of miles away, and (to her disappointment) far from the action. The scene is particularly effective because Demme breaks the cinematic convention, therefore leaving us surprised and at the edge of our seat with worry for Claris.[87] But while some of each conventional narrative film can be processed in terms of those previously learned schemata, the film's effectiveness seems to arise, as in the case of *The Silence of the Lambs*, from the particular and unique moments -- those which were neither anticipated, nor called for, but that could have been justified in the context of the plot. In order to account for these moments, let me now turn to a discussion of mental imagery.

---

[84] W. G. Chase & H.A. Simon, "Perception in Chess" Cognitive Psychology, 1973: 4, pp. 55-81.

[85] Ibid.

---

[86] See appendix #5.

There are two ways in which a mental image can be formed: one is during the time of perception, while retaining an online input (i.e., seeing a picture, or the environment), and the other is by activating information which is stored in long term memory. This second case involves "seeing" in the absence of an immediate sensory input. Stephen Kosslyn claims that imagery is used when

> (1) The information to be remembered is a subtle visual property; (2) the property has not been explicitly considered previously (and hence labeled); and (3) the property cannot easily be deduced from other stored information (for example from an information about the general category to which the object belongs).[88]

This observation is important for narrative comprehension, as so much of it happens in retrospect, by activating memory. It is interesting therefore to note that sometimes the categorical condensation is not enough, and one needs to go back and search for a particular image, which will then be able to provide the information needed. In *The Silence of the Lambs* for instance, Claris is certain she

---

[87] A fuller discussion of this scene can be found in chapter 3.

[88] Stephen Kosslyn, op. cit. p.268.

has arrived at the killer's house once she sees bugs in the kitchen. A close-up of the bug helps her, and us, recall previous information about bugs that came up during her investigation, and we need no more explanation for why she immediately pulls out her gun. In structural narratology this moment would be classified as a paradigmatic moment, where our memory is activated by some cue, and at once we have a reference to something that happened earlier. A paradigmatic moment condenses the linearity of the plot, and brings forth the layering of information relevant to the construction of the narrative. In the case of Claris, the paradigmatic moment was excited by an on-line image, which then activated the retrieval of a mental image from memory. The striking fact about the nature of this memory retrieval is that it is extremely local and particular. We don't just recall the category "bug" but we recall the *exact* bug Claris saw previously, in the context of which she saw it. How can such a specific image arise if long term memory requires an abstraction and condensation? As mentioned before,

> two means of representation have been proposed for mental images, one that confers a special status on images and one that treats them as no different in kind from the representations of linguistic meaning. The two alternatives are called depictive and propositional representation.[89]

These types of representations use different formats, or different kinds of codes. Each system has a syntax, which includes the symbols and the set of rules that enable combining these symbols. The semantics is determined by how meaning is conveyed by symbols and combination of symbols, and the content is the specific information that is conveyed. In the case of propositional representations the symbols in the syntax belong to a variety of form classes, which include entities, relations, properties, and logical relations. The semantics is determined by arbitrarily attributing meaning to individual symbols, which requires the existence of a lexicon (much like words in natural languages). A propositional representation is also abstract, as it can refer to non-picturable entities (such as headaches, or feelings), and it can refer to classes of objects. Depictive representations, on the other hand, are not abstract, they represent individual cases (rather than categories) and they cannot refer to non-picturable concepts. The symbols are composed from two form classes: points and empty space, and there are no defined rules of combination, that is, there is no visual syntax determining how points should/could be combined. The relationship between a depictive representation and what it stands for is not arbitrary, but based on visual resemblance.

---

[89] Stephen Kosslyn, op. cit. p.280.

In the last two decades there has been an on-going debate whether mental images are depictive or propositional, and experimental support has been provided for both models. Recent years have shown that there are regions of the visual cortex that are topographically organized, and that retain the spatial structure of the retina. Connections from this area of the cortex do not move only downstream (from sense experience to be processed by higher cognitive mechanisms), but rather run both ways, that is, from previously memory stored images back to areas of the brain where they are activated and used.[90] As a result Kosslyn concludes that

---

> These facts are consistent with the notion that visual memories are stored in an abstract (propositional?) format and that an image is formed in order to make accessible information about the local geometry of a shape. An image is formed, presumably, by using the backwards connections that run from the areas involved in visual memory to (at least some of) the areas that are topographically organized. The image would make accessible spatial information that was only implicit in long term memory representation. If so, then image representations would be depictive in the strongest sense of the term: they would be patterns in a physical, and also functional, space. Hence they would literally be "pictures in the head."[91]

At this point in time there is not only behavioral evidence for these findings, but also neurological support for the active role played by topographically organized parts of the brain in imagery. It seems that when we recall images we use both propositionally and depictively encoded visual information. When we recall, together with Claris, the images of the bugs from early on in the film, we retrieve a propositional memory, which gets represented depictively, so that we can actually see the bug we saw earlier, and not "a bug" in general.

Summary

In this chapter I have shown that images are perceived and cognitively processed in significantly different (and sometimes even dialectically opposed) ways to language.

---

[90] Ibid. p.290.

[91] Ibid. pp. 290-291.

To sum-up we may say that language perception and cognition operates by a serial processing that takes abstract symbols, matches them to a pre-learned lexicon, and organizes them in operational logical relations. These decoded linguistic signs are stored in memory as prototypical and propositional sets, and are retrieved as such when high-order cognitive operations are active.

Visual perception, on the other hand, works both holistically on faces, and in decomposition into parts of objects. It is transmitted both as analog representations and as categorical (i.e., prototypical) and propositional ones. It is, however, stored in long term memory as an abstract propositional set of symbols, but, when activated by memory, it is sometimes re-translated into a depictive and analog image. The implication is that when we use visual memories, much as at the time of perception, we are using both pre-attentive and focal attentive practices, and we use automatic, (or parallel) as well as focal attentive (or serial) searches. We use grouping and holistic practices to process parts of the image, while we decompose other parts into geons.

It may be safe to assume that in the case of film, general (or the background) spaces are processed automatically and holistically, so that attention can be focused on action, which is more linear and causal by its nature, and is therefore subject to serial processing. I would speculate that close-up shots of objects will be processed serially as well, as they are shots emphasized by the director, as if the filmmaker says: "Pay attention, this is important." Some of these different visual materials in the frame, their processing, and their relations to one another and to the plot, will be examined in following chapters.

Whether the actual breakdown of the frame is holistic or serial at the time of perception, it is clear though that unlike language, which is processed propositionally, images are cognitively processed both propositionally and depictively. We can now return to Dretske's notion of meaningful perception as a perception that requires recognition, categorization and computation. And while it is now clear that for visual material to be stored in long-term memory it needs to undergo these processes, it seems that at the time we use visual data for cognitive activities (like constructing a narrative), we do not necessarily use encoded, categorized visual messages. In other words, at the time that visual material is being cognitively operated on by high-order mechanisms, it is not necessarily either abstract or propositional, and at least part of the visual material can be processed as holistic and/or analog. Language, on the other hand, may invoke mental imagery, but is processed in abstract, categorized, symbolic networks. The implication of this data is that at the time of the construction of the narrative we use both information that is encoded into logical chains (propositions), and information that is not categorized or abstracted (aspects of the image track). And since both kinds of data (propositional and depictive) are important for the understanding of the narrative, we may declare both as meaningful perceptions. In the next chapters I will show that a semiotic approach to film treats it primarily as a linguistic medium. I shall claim that we need to treat cinema as a complex medium that employs both linguistic and image-based information system, and I will explore further the implications of the dynamic and multiple processing activities of these information systems to the narration and comprehension of film.

Chapter 3

A Cognitive Approach to Film Narration

Introduction

In the first chapter of the dissertation I identified a problem in film narratology, namely, that it relies heavily on literary narratology, thus not accounting for the specific nature of the filmic medium, and its implications for cinematic narratology. In particular, I looked at two literary narratology terms – enunciation and focalization, and how they were imported into discussion of film narration and point of view in simplistic and problematic ways. At the end of the chapter I claimed that in order to have a complete account of film narration, one needs to account for the ways in which the audience actually perceives and processes filmic information: i.e., dialogue, images, sound effects and music. In the second chapter I reviewed cognitive science research on visual and verbal perception, and have shown how such information can help us understand how audiences perceive and record in memory particular moments in film.

In this chapter I will propose a cognitive model of film narration. I will first show how a cognitive reading can bypass the limitations of a semiotic/structuralist narrative model, and then discuss the benefits of a cognitive model. With the aid of the conclusions of the second chapter I will propose a model of film narration, and particularly of point of view editing. At the end of the chapter I will analyze the film *Rambling Rose* according to this cognitive model of narration.

Semiotics

Film narratology, narrative analysis, or the theoretical study of film narrative, is a semiotic based theory, a theory that examines cinema as first and foremost a signifying system. Early in the century, theorists like Eisenstein, Tynianov, Shklovsky, and Eikhenbaum set some preliminary terminology in a formalist discussion of cinema, attempting a close material analysis of cinema's basic building blocks.[1] Christian Metz, Peter Wollen, and Todorov have advanced the formalist discussion to a larger structuralist framework, one that (following Lévi-Strauss, Propp and Jakobson) focused more on plots, functions, and underlying structures. The primary concern of film semiotics was (and to a degree still is) to identify structural mechanisms at work in film, such as plot structures, characters, modes of narration, temporality, etc. While structuralist in nature, semiotics in its turn has been both used as a tool and criticized in psychoanalytic film theory as well as in feminist and Marxist film criticism.

Film semioticians have provided an array of useful terminology and methods to analyze different aspects of film narratology, but all of those terms are based on basic distinctions with regards to the cinematic sign. Film semiotics relies on ground breaking work that was done by two turn of the century semioticians, Ferdinande de Saussure and Charles Peirce. In film theory, the most comprehensive description of their work and its implications for cinema was provided by Christian Metz. But I find Metz's description to be somewhat lacking, at times confusing, and in the end he draws some flatly wrong conclusions. And given that most other film semioticians draw on Metz and his terminology (even when they argue with him), the flaws of his semiotic interpretation have infiltrated the whole debate. I shall therefore briefly describe the work of de Saussure and Peirce, then its cinematic importation, and finally criticize it with the aid of cognitive science.

---

[1] See for instance: Sergei Eisenstein, Film Form: Essays in Film theory, trans. By Jay Leyda, New York: Harcourt, Brace & World, 1949; Boris Eikhenbaum, The Poetics of Cinema, trans. By Richard Taylor, Oxford: RPT Publications, 1982.

Already when discussing language perception in the last chapter I mentioned the work of the semioticians Ferdinande de Saussure and Charles Peirce. Both claimed that thought cannot be separated from the language in which it has been expressed. Saussure, a turn of the century linguist, started thinking about language as a system of communicative signs, such as deaf and mute sign languages, symbolic rites, forms of politeness, road signals, etc.[2] All these communicative systems share a structure by which each linguistic sign is comprised of a concept and a sound-image.[3] The concept of a pet feline or the signified is expressed by a signifier -- the vocal sounds c-a-t, or by an image of that same animal. Saussure claimed that the means of expression in a society are based on a convention, or a cultural agreement. Natural languages can be seen as a special case of a semiological system, where the sign is not only conventional but abstract and arbitrary. In other words, there is no necessary connection between the sound pattern d-o-g and the animal it designates. The connection between the signifier and signified is made purely by means of a group's (in this case the English-speaking society) implicit agreement.[4] In order to use natural languages one needs to learn the conventions that govern that communication system; i.e., its lexicon, syntax, and pragmatics of use. De Saussure went on to say:

---

[2] Ferdinande de Saussure, "The Object of Study," in David Lodge, editor, Modern Criticism and Theory, London and New York: Longman, 1988, p.8.

[3] Ferdinande de Saussure, "Nature of the Linguistic Sign," Ibid. p. 11.

> Signs that are wholly arbitrary realize better than others the ideal of the semiological process; that is why language, the most complex and universal of all systems of expression, is also the most characteristic; in this sense linguistics can become the master-pattern for all branches of semiology although language is only one particular semiological system.[5]

What is stated here is that the more highly coded a sign system is -- the more abstract and conventional -- the more effective it is as a communication system. And therefore a hierarchy of sign systems is created, in which natural languages are better examples of the ideal semiological system. De Saussure, being a linguist, was mostly interested in natural languages. Charles Peirce, the American pragmatist philosopher, was interested in the relations between language, thinking and the mind. He set out to attack the Cartesian model in which thinking is an immediate perception of ideas within the mind or soul.[6] For Peirce, much like for Saussure, there is no thinking without signs, but while Saussure was interested in language as a closed system of differences, Peirce is interested in how we obtain knowledge about the world. Peirce concluded that we have no power of introspection, but all knowledge of the internal world is derived by hypothetical reasoning from our knowledge of external facts. When debating the question of whether we can think without signs, Peirce claimed:

---

[4] The implicit agreement, or development of a conventional system that connects signifiers to signifieds is subject to much research by linguists and still pose an array of interesting problems, since it is not an explicit process.

[5] Ferdinande de Saussure, "Nature of the Linguistic Sign," in H. Adams and L. Searle (eds.) Critical theory Since 1965, Tallahassee: Florida State University Press, 1986, pp. 647-648.

> If we seek the light of external facts, the only cases of thought which we can find are of thought in signs. Plainly, no other thought can be evidenced by external facts. But we have seen that only by external facts can thought be known at all. The only thought, then, which can possibly be cognized is thought in signs. But thought which cannot be cognized does not exist. All thought, therefore, must necessarily be in signs.[7]

Peirce defined three kinds of signs and their relations to dynamic objects and our knowledge of these objects.[8] The first, the iconic sign, is based on a relationship of visual resemblance between the signifier and the signified. A drawing of tracks on a road sign is an iconic signifier for upcoming train tracks, and a photograph of the Eiffel tower is a signifier for that material construct in Paris. Naturally, most film images are iconic signs, and given that the technology used to create these images is photographic, the level of resemblance between signifier and signified is high. In fact, one may claim that the signifier is an unmodified representation of what was placed in front of the camera, or the signified. Even though focal length, framing, and other technical constraints of the camera may not replicate the way our eyes see, the fact remains that the objects that were seen by the camera (through the camera lens) at the time of taking the photograph, are seen by the viewers of the photograph in the same spatial arrangement, same framing and same lighting. The photographic signifier is therefore sometimes assumed to be interchangeable with, or identical in appearance to the photographed signified.

---

[6] Charles Peirce, "Questions Concerning Certain Faculties Claimed for Man," in Collected Papers Vol. V, ed. Charles Hartshorne and Paul Weiss, Cambridge, MA: Harvard University Press, 1931. Pp. 135-155.

[7] Ibid. p.151.

[8] Charles Peirce, "The Icon, Index, and Symbol," in Collected Papers Vol. II, ed. Charles Hartshorne and Paul Weiss, Cambridge, MA: Harvard University Press, 1931. Pp. 156-173.

An indexical sign is based on an existential bond or relationship between the signifier and the signified. Smoke signifies fire, and a bullet hole signifies the fact that a gun was fired. There is a causal and historical (temporal or spatial) bond between the signifier and the signified, hence the existential component. Sound effects that do not originate from the visual field can be seen as a good example of indexical signs.[9] A knock on an (invisible to the audience) door indicates both an expansion of the visual space beyond the frame lines, as well as the (dramatic) arrival of a character. Sound effects which accompany interior scenes indicate spatial placement (urban, rural, etc.) thus indexing a larger spatial frame of reference. On a different level altogether, a photographic image also functions as an indexical sign -- in this case, a historical one. Once (early 1970s), an actor (Martin Sheen), stood in front of a rolling film camera in south east Asia. Every time I watch the film (*Apocalypse Now*, Coppola, 1978) the images stand in indexical relations to the event of making that film.

---

[9] Sound effects are often recorded in a folly studio, where many materials are used to create a sound similar to the desired effect. That is, the signifier is produced not by the signified object, but by something else. And while it may seem that the index in this case is not of what they seem to be. It is important to remember in this respect, that the signified is not the object itself, but the part of the sign which is the idea of the object. That is, a folly sound of horse hooves which was produced by wood on tile, is not a lying

A symbolic sign is defined as a sign in which the signifier's ability to represent the signified is dependent upon a rule. Much like Saussure's description of the linguistic sign, the symbolic sign is abstract (chains of letters represent objects), arbitrary (there is no necessary conditions that determine that d-o-g, and not c-a-t should stand for the signified canine that barks), and it is conventional. Natural languages are the best example of symbolic based sign system. While some conventions have developed in cinema -- such as the use of certain musical arrangements to signify particular moods, or the use of lighting to create aura behind the main (good) character, or low angle shots to convey fear/submission – these are not arbitrary, or abstract conventions, but have developed in accordance with organic cultural conventions. It is hard therefore to imagine a symbolic sign system other than the linguistic one to operate in cinema.

---

indexical signifier, since the signified is the idea (horses galloping), not the actual thing which created it (wood on tile).

Semiotic Film Theory

Peirce's neat division into iconic, indexical and symbolic sign systems was very attractive to semioticians in film theory. While the Russian formalists attempted to explain cinema as analogous to language, and therefore spent much time and effort in determining the basic film unit (the equivalent of a word in natural languages) as the shot, and a montage as a phrase, Peirce's terminology enables the critics to decode individual signs based on the embodying medium. Christian Metz in particular tried to define whether cinema can be seen as a language, or as employing a language system, and after a thorough semiotic comparison between verbal languages and cinematic ones he determines:

> First of all, the shot, through its semantic content [. . .] is closer, all things considered, to a sentence than a word. An image shows a man walking down the street: It is equivalent to the sentence "A man is walking down the street." The equivalence is rough, to be sure, and there would be much to say about it; however the same filmic image corresponds even less to the word "man" or the word" walk" or the word "street," and less still to the article "the" or to the zero degree morpheme of the verb "walks."[10]

After great deliberation Metz concludes that film is like a language, since it has an expressive-communicative content, but it has no natural language-like system of rules, particularly, no rules of denotation that define a specific relationship between a signifier and a signified in the

---

symbolic fashion.[11] Semiotics here is used both to see how cinema is and is not behaving like natural languages, as well as to define some initial concepts about film comprehension and interpretation.

Semiotics, and particularly the Peirceian view, were useful for non-structural purposes as well. Since the coming of sound, film theorists have been interested in the illusion of reality, or the impression of reality that a film manufactures, and in its effects on the audience. Bazin hailed realism as a style,[12] but feminists, Marxists and other political readers of film have been concerned with the production of the impression of realism and feared its impact – particularly that of neutralizing ideology – on audiences.[13] Explaining the cinematic image as an iconic sign is attractive to the "illusionists" since it implies a certain immediateness about that signifying system, a very tight bond between the signifier and the signified, and therefore a difficulty (for the audience) in recognizing that what they are watching has been intentionally produced and constructed for their entertainment. When comparing the relations of the signifier and signified in natural languages and in cinema, Metz claims that in the case of language there is a distance between the "content" and "expression." In contrast,

---

[10] Christian Metz, Film Language: A Semiotics of the Cinema, Chicago, IL: The University of Chicago Press, 1974. pp. 66-67.

[11] Ibid. Ch. 3: "The Cinema: Language, or Language System?"

[12] André Bazin, What Is Cinema Vol. I & II, Berkeley, University of California Press, 1967.

[13] A good account of the theorists who attempt to decode the impression or illusion of cinematic realism can be found in Noël Carroll, Mystifying Movies: Fads & Fallacies in Contemporary Film Theory, New York: Columbia University Press, 1988, chapter 3, "The Cinematic Image."

> In the cinema the distance is too short. The signifier is an image, the significate is what the image represents. Furthermore, the fidelity of the photographic process, which gives the image particular verisimilitude, and the psychological mechanisms of participation, which ensure the famous "impression of reality," shorten the distance even more – so that it is impossible to break up the signifier without getting isomorphic segments of the significate.[14]

Metz then proceeds to meditate more about the nature of the cinematic sign, and as he is fascinated by the expressive and connotative value of the image itself, he slowly veers away from its function as referring to an object. Metz concludes:

> There are many characteristics to the filmic image that distinguish it from the preferred form of signs – which is arbitrary, conventional, and codified. These are the consequences of the fact that from the very first an image is not the indication of something *other than itself,* but the pseudopresence of the thing it contains.[15]

For Metz, then, at a very basic level, the cinematic image is an iconic sign where the signifier refers to itself, rather than to the signified, although it gives the impression or the pseudopresence of the signified. This approach to the cinematic sign (which is very similar to Derrida's "free floating signifiers") was welcomed by theorists who attempted to decode the ideological mechanisms of the impression of reality film manufactures.

---

[14] Christian Metz, op. cit., pp. 62-63.

Bill Nichols, for instance, wholeheartedly accepts this semiotic approach, and when discussing the ideology in the image he claims:

> Since images bear an analogous or iconic relationship on their referent (a relationship of resemblance), it is easy to confuse the realms of the image and the physical world by treating the image as a transparent window (especially the photographic image), or by treating the physical world idealistically by assuming that something like its essence has been transferred or reproduced in the image. Many films employing realist styles encourage such a confusion, and yet it is essential to remember that a film is not reality any more than an image is what it re-presents. [. . .] We might even say, metaphorically, that realist images are an objectification, or projection of the normal perceptual process. What our nervous system initially encountered as unorganized sensory input is now encountered as the organized or signifying output of these objectification, or images.[16]

Much like Metz, Nichols here assumes an erasure of the signified, or its reconstruction in the mind of the audience in terms of an ideologically controlling signifier. Secondly, the audience, *à la* Nichols, is led to exchange signifier with signified because of the mimetic, verisimilitude quality of the photographic image. In return, this approach enabled many contemporary theorists (labeled by Noël Carroll the "Psychosemiotic

---

[15] Ibid. p. 76. My emphasis.

[16] Bill Nichols Ideology and the Image, Bloomington: Indiana University press, 1981, p.24.

Marxists"[17]) to talk about images not as representations, but as constructs, which refer to other filmic constructs. And although at some basic level it makes sense to acknowledge that while watching a film we are presented with an artistic construct, such a direction takes us away from understanding how we infer meaning from visual signs, and how we construct interpretations of actual films on that basis. In other words, as audience, we are always aware that we are watching a man-made film that was produced by a filmmaker and a professional crew. But while we are watching a film, we are perceiving and processing visual information, which we understand to stand in relation of (at least partial) resemblance to the visual, photographed world. That is, at the time of interpretation we are concerned with the content of the images, and not with the fact -- which we readily accept about the nature of film – that they are constructed images, or signs.

An additional problem arises when Metz extends his observations about the image to the whole filmic experience. Metz claims: "the cinema begins where ordinary language ends."[18] The problem with such a claim is that it assumes the superiority of the image over other signifying systems. In other words, Metz here claims that what is true of the iconic sign is true for the whole cinematic text, thus ignoring the existence and functioning of indexical and symbolic signs in cinema. In the previous chapter I have shown that the perception of images and of natural languages occurs in somewhat different cognitive manners. I believe that attempting to lump all cinematic signifying systems together into a unified "cinematic sign" is not only theoretically reductive, but as a result, prevents us from an adequate explanation of the complexity of the film experience.

---

[17] Noël Carroll, Mystifying Movies: Fads & Fallacies in Contemporary Film Theory, New York: Columbia University Press, 1988, p. 107.

[18] Christian Metz. Op. cit. p. 81. Here Metz is talking about the denotative nature of language and the connotative/expressive nature of the image. I shall not go into the whole distinction as it is irrelevant to the discussion I am pursuing, but I would like to emphasize how easily Metz is able to ignore the linguistic aspect of cinema.

Peter Wollen, when discussing semiotics and cinema is much more inclusive of other sign systems, but still quite hierarchical:

> In the cinema, it is quite clear, indexical and iconic aspects are by far the most powerful. The symbolic is limited and secondary. But from the early days of film there has been a persistent, though understandable, tendency to exaggerate the importance of analogies with verbal language. The main reason for this, there seems little doubt, has been the desire to validate cinema as an art.[19]

Wollen here is willing to address cinema as employing all three modes of signifying systems, and he criticizes the theoretical practice of valuing a linguistic analysis. Furthermore, Wollen asserts:

> The great merit of Peirce's analysis of signs is that he did not see the different aspects as mutually exclusive. Unlike Saussure he did not show any particular prejudice in favour of one or the other. Indeed, he wanted a logic and a rhetoric which would be based on all three aspects. It is only by considering the interaction of the

---

[19] Peter Wollen, Signs and Meaning in the Cinema, Bloomington: Indiana University Press, 1972, p.140.

> three different dimensions of the cinema that we can understand its aesthetic effect.[20]

But Wollen too, thinks of the iconic sign as "shifting and elusive", as non-conceptual, and in short, as placed in dialectical opposition to the symbolic.[21]

But Metz's and Wollen's views on the nature of the iconic sign are not grounded in semiotics, and actually are quite contradictory to Peirce's and de Saussure's visions. In his work on phenomenology Peirce defined three states of being in the world and relating to it.[22] Firstness designates a monadic reality, where no relations to others exist. Secondness is the force that demarcates one thing from another (either by brute force, or by self identity), thus creating a relation, or a correlate. Thirdness implies a power of mediation that brings the dyadic relation (between firstness and secondness) to a higher form of rationality. This thirdness is a representational relation that is intelligible and manifests law-like regularity. The signifier thus is a first, which stands in a relation to a second, or its signified object. The nature of the relations between the signifier and the signified is determined by a rule, which Peirce called the interpretant, a third, a convention, or better, a social practice. On the face of the matter it seems like the iconic sign has a relation of firstness, an indexical of secondness, and a symbolic of thirdness. This attitude would, of course strengthen the illusionists claim for the power of the cinematic image; an icon as a first refers only to itself, or, better, *is* just itself. But Peirce claimed that all sign systems are based on a relation of thirdness, that is, that all signs are regulated by a law, or a conventional representational system.

---

[20] Ibid. p.141.

[21] Ibid. p.152.

[22] Charles Peirce, "The Categories in Detail," in Collected Papers Vol. I, ed. Charles Hartshorne and Paul Weiss, Cambridge, MA: Harvard University Press, 1931. Pp. 148-180.

> A *Sign*, or *Representamen* [=signifier], is a First which stands in such a genuine triadic relation to a Second, called its *Object* [=signified], as to be capable of determining a Third, called its *Interpretant* [. . .].[23]

An icon, then, is as regulated by a law as a symbol or an index. The presence of the interpretant, an agency that determines the relations between the signifier and the signified, makes it clear that one cannot abandon the signified at all, as Metz is trying to do, or attribute a conceptual basis to symbolic signs but not to iconic, as Wollen is claiming. At the heart of semiotics stands a social practice that ties the signifier and the signified to form a sign, or as de Saussure puts it:

> In fact, every means of expression used in society is based, in principle, on collective behaviour or – what amounts to the same thing – on convention.[24]

If one wants to use semiotics as an explanation of cinematic signification systems, one needs to be able to account for the cultural convention that govern iconic representation.[25] But Metz's and Wollen's omissions of the convention, or the interpretant, are understandable, since the photographic/cinematic image resist that kind of description. Our experience of viewing does seem automatic, immediate, and un-coded. It is very difficult to explain the cultural codes that are at work in viewing a specific photographic image. The idea that images do not duplicate reality, but somehow, through ideological manipulation, give the impression that they duplicate reality, can only work as a global account of film production, not as an account of specific images. When coming to decode individual signs (say for instance an image of a gun), one is at loss as how to proceed with an analysis of an ideological cinematic iconic representation. Moreover, while the theoretician accepts that an iconic sign is not a natural object (does not re-present reality), they attribute that belief to the masses. Or in Carroll's words:

---

[23] Charles Peirce, "The Icon, Index, and Symbol," op. cit. p. 156. The inserted brackets are my addition.

[24] Ferdinande de Saussure, "The Nature of the Linguistic Sign," op. cit. P.647.

[25] For the conventional aspects of "natural" representation in art see also E. H. Gombrich Art and Illusion, Princeton: Princeton University Press, 1960.

> Put crudely, psychosemiotic Marxists reject Bazin's theory of the ontology of film for themselves but, so to speak, attribute a belief in it to ordinary spectators.[26]

So even though film semioticians are willing to accept the constructed and conventional nature of the sign, they believe that regular audiences watch and process cinema as if it was a natural and unmediated reproduction of the world. And if audiences could actually successfully interpret films without recognizing them as signifying systems, what is the merit of a semiotic analysis of film?

---

[26] Noël Carroll, op. cit. P. 114. As one example, one may think of the ongoing debate on the influence of violent cinema on the behavior of audiences. The critics usually attribute a sense of naiveté, or unawareness to the audiences, particularly if they are young.

I would like to propose two answers to this question. On a global level, as I mentioned earlier, I believe that the audience is always *a priori* aware that they are watching a symbolic and cultural construct – a film. In the case of fiction cinema it is clear that a tale is being told, a story being visualized, a dramatic construct is designed from the ground up. But even in the case of a documentary, which is supposed to be making some truth claims about the real world, audiences are aware that the representation is mediated through the eyes of the filmmaker -- that framing was chosen, certain questions asked, and editing was done -- and thus that there is a creative force behind the text they are watching.[27] On that global level, a semiotic reading of a film is appropriate. In the case of fiction film, which is the focus of this thesis, it is a semiotic analysis of the narrative, of the signifying systems used to convey the story or the plot of the film. In structural terminology it is an analysis of how the syuzhet leads audiences to construct the fabula.

On a second level, the level of the communicative systems film employs to tell its story -- or in Bordwell's terminology the film's *style* [28] -- one needs to be careful. In the following pages I will show that certain aspects of visual perception cannot be explained in semiotic terms. That is, visual perception is not based on decoding a conventionalized sign. At the same time, the storage of visual information in long-term memory probably does involve propositional codification that is very similar to a symbolic sign system. Using cognitive science I will show that on the level of perception of the syuzhet (through the style) one needs to be careful of employing semiotic interpretation, but on the level of the construction of the fabula, there is more room for such an analysis. In other words, while high-order cognitive operations function similarly to language, and can therefore be evaluated by semiotic standards, bottom up perception should be analyzed according to the separate channels of information (i.e., images and verbal information).

---

[27] And even if one wanted to discuss surveillance camera material, the camera has been placed in a certain place, set to a certain lens, with a certain focal length, by a human agency.

While Metz is willing to abolish the iconic signified in order to deal with the apparent lack of distance between signifier and signified, I would like to suggest that we, as viewers, focus on the impression that there is only a signified, even though we are aware that we are watching signs which have been produced, and organized for our consumption. That is, at the time of viewing we are knowingly and willingly suspend our disbelief, or our knowledge of artificiality, in order to focus on the interpretation of signified events, rather than on the signifier. And if we understand why the images of a film give us this unmediated, non-produced impression, we can understand the low-level perception of the visual aspect of a film, and how it leads to interpretation. But before discussing perception, I need to reject one other myth promoted by Metz, the myth of cinema as a language.

The semiotic approach encouraged the structuralists to explain our competence in understanding non-linguistic representations by appeal to

---

[28] David Bordwell, Narration in the Fiction Film, The University of Wisconsin Press, 1985, p. 50.

linguistic mental operations. After all, while the semiotician attempts to examine all signifying systems, the one best studied, best understood, and according to some, best at communicating, is the natural language system. And while recognizing that cinema did not operate by a governing system of rules, like natural languages do, Metz insisted that since cinema is communicative it is a language. But cognitive science indicates otherwise. Following Chomsky, Gregory Currie points out that communicating with the use of natural languages involves two aspects: productivity and conventionality.

> Productivity means that an unlimited number of sentences of English can be uttered and comprehended; in fact many of the sentences we utter and comprehend every day have never been uttered, and so have never been comprehended, before. [. . .] English is conventional in that what words and sentences of English mean is determined [. . .] by adventitious uniformities of practice adhered to because they aid communication.[29]

Much in agreement with the semioticians, the idea of conventionality accounts for the rules of denotation between signifier and signified. But language competence is not generated just by learning those conventions. Meaning is a result of a practice (productivity) in which an agent produces utterances by combining words (signifiers) to form sentences. While conventionality determines that language has to be learned, "productivity precludes the language being learned sentence by sentence, since [. . .]

---

[29] Gregory Currie, Image and Mind: Film, Philosophy, and Cognitive Science, New York, NY: Cambridge University Press, 1995, pp. 120-121.

speakers understand sentences to which they had no previous exposure."[30] The individual words, in Currie's account, become meaning atoms, and since the meaning of a sentence is dependent on the lexical meaning of individual words, meaning in natural language is acontextual.

Cinema, in comparison to natural language, exhibits productivity but not conventionality. While watching a film we are generally shown many images and image combinations which we have never seen before, and which we have no problem understanding. But unlike language, the cinematic image is not governed by convention in any way similar to the function of conventions in the production of literal meaning in language. Currie claims:

> There are no atoms of meaning for cinematic images; every temporal and spatial part of the image is meaningful down to the limits of visual discriminability.[31]

Currie is not claiming that cinema is not influenced by conventions such as eye-match cuts, or the conventions associated with fades or dissolve. His claim is that the cinematic image is not governed by acontextual, molecular, convention in the way words in natural language do, and therefore that the image cannot be purely conventionalized to discrete units of meanings like words do. While cinematic metaphors and conventions have emerged, there is no cinematic equivalent to the literal conventional system of denotation in natural languages. Currie is also not claiming that language functions only by conventional rules. Context and interpretive practices are important, but cannot function outside or without the convention or the lexicon. As for the generation of meaning in cinema Currie concludes:

---

[30] Ibid. pp. 121-122.

> Being without meaning atoms, and therefore nonrecursive, and being at the same time productive, the meaning of cinematic images cannot be conventional. It is natural meaning. With images, productivity is natural generativity, and we explain that in terms of natural recognitional capacities.[32]

Currie's account is significant, since it rejects the idea that we can understand cinema in linguistic or semiotic terms. Images are not perceived and understood as conventional signs, but as natural images. While the semiotician promotes the idea of *seeing as,* the cognitivist uses the term *seeing in.*[33] The idea here is that when seeing an image of an object A, we recognize *in* the picture the object A. How we *see in* rather than *see as* is the topic of the next section.

Style – or Bottom Up Perception

In a recent book which provides an ecological approach to film theory, Joseph Anderson relies on J.J. Gibson's theory of direct perception to explain film perception.[34] Gibson's approach, as mentioned in the

---

[31] Ibid., p. 130.

[32] Ibid., p.131.

[33] The term was coined by Richard Wollheim in Art and Its Objects, New York, NY: Cambridge University Press, 1980.

[34] Joseph D. Anderson, The Reality of Illusion: An Ecological Approach to Cognitive Film Theory, Carbondale and Edwardsville: Southern Illinois University Press, 1996.

second chapter, is that perception is direct – that is, it need not be mediated, or interpreted by some inferential process. One of the main objections to Gibson, presented by cognitive scientists like Fodor, Pylyshyn, and others, is that Gibson's theory cannot explain visual illusions, as there is no room in Gibson's theory for non-veridical perceptions.[35] But a computational correction of Gibson's theory can explain the misperception of illusions.[36] According to such theory, the neural system, which is based on both electric and chemical transformations, is very dependable but slow. In order to speed up the perceptual processes, some assumptions (about closure of objects in two dimensional representations, overlapping, and movement) are made so as to create "shortcuts" in visual processing. While useful in most occasions, these shortcuts can mislead us to judge visual information in non-veridical ways. And film, as we have seen, is a representation, that is a man-made construct, which, by virtue of its photographic technological base, creates an "illusion or functions as a surrogate reality."[37] Anderson makes a strong case for examining the perception of film in much the same ways we examine the perception of the world, as a sub category of illusionist perceptions, or a case of "ancient biology interfacing with recent technology."[38] Throughout the book Anderson provides much evidence showing that cinema has evolved to better use our natural perceptual devices, and that this evolution can account for the realistic impression of film. To name just a few of Anderson's examples, we can consider the development of the standard film (or video) speed, and the technology that was advanced to eliminate flicker. While film speed was unfixed before stabilizing on the 24 frames per second rate, the problem of flicker was only fully resolved once the multi-bladed shutter was created.[39] Both film speed and shutter flicker have been standardized once they created a sense of continuous, flowing motion, rather than the jumpiness of intermittent frame movement in front of the projector lens at the turn of the century. Similarly, Anderson examines the phenomenon of linear convergence, or perspective. Theorists like Baudry and Panofsky have claimed that Renaissance perspective is a cultural convention, and thus photography and cinema, by use of the photographic lens, inherently impose a Western perspective on its viewers.[40] But Anderson shows that the representation of the convergence of light in a perspectival manner, (which was invented during the Renaissance), is merely based on our biological (in-the-eye) lens, which already employs perspective as means of discerning depth and size.[41] Noël Carroll, too, points out that convergence perspective was not invented by the Renaissance, but discovered during that period and that, "the perspective system is more accurate, transculturally, in terms of affording certain spatial information than any other mimetic pictorial system."[42] Anderson's other example of how film evolved to resemble the perceived visual world, and therefore interface better with our perceptual capacities, is the issue of synchronicity. Research shows that a baby's attention span on objects is much longer when those objects emanate synchronous sound.[43] Once hearing sounds, babies have been attempting to visually locate the source of the sound, and if successful at that, they classify the auditory and visual signals as one event. Synchrony in our environment is important for perceptual analysis, since it indicates that some cross module (hearing/seeing) activity occurs already at very initial developmental stages of perception. The early children's search for synchronicity also indicates that our perceptual mechanisms are not passive, but actually actively seeking connections, or affordances, to use Gibson's terminology.[44] Synchronous film technology, which is at the base of realist classical cinema, relies on these very basic perceptual tenets. Anderson shows that even music is synchronized a great deal to the visual events on the screen, and that this synchronicity helps the audience connect the non-diegetic music to the narrative world of the film.[45] These examples and others show that realist cinema is most successful when adhering to very basic perceptual mechanisms, ones that have developed not for the sake of consuming cinema, but in order to survive and better cope in our environment.

---

[35] J. Fodor and Z.W. Pylyshyn, "How Direct is Visual perception? Some reflections on Gibson's 'Ecological Approach'," Cognition 9, 1981.

[36] Here I am following Anderson's description of the work of David Marr and others in the first chapter "Toward an Ecology of Cinema," op. cit. Pp. 28-35.

[37] Ibid. P. 19.

[38] Ibid. p.28

---

[39] Ibid. p. 54-61.

[40] See Jean Louis Baudry, "Ideological effects of the Basic Cinematographic Apparatus," Film Quarterly 18, 2 (winter 1974/75) pp. 39-47, and Erwin Panofsky "Style and Medium in the Motion Pictures," Critique 1, no, 3, January- February, 1947.

[41] Joseph Anderson, op. cit. P. 72.

---

[42] Noël Carroll, op. cit. P. 130.

[43] Joseph Anderson, op. cit. . pp. 82-83. Anderson quotes the work of cognitive scientists like Held, O'connor and Hermelin.

[44] Ibid. pp. 49-53.

Anderson claims:

> . . . it is the fact that the perceptual systems go through the same computational procedures whether confronted with the real world or with synthesized shadows and sounds that allows for the existence of cinema.[46]

While not ignoring the specificities of the cinematic medium, Anderson suggests that in order to understand the perception of cinema we need to understand our perception of the world, and that such a theory of perception would generate a theory of cinema which is positioned in opposition to linguistic or politically based theories of cinema.[47] Anderson's merit is in showing that at the initial levels of perception of the filmic visual material, our cognitive processes operate with film in much the same way that they do with the real world, i.e., they perceive and process bottom–up information without categorizing and employing high level cognitive activities. While language can be perceived and processed only with the aid of pre-, or sometimes co-learned categories of denotation, images are perceived and understood as is. Anderson's argument pertains to perception only, that is, the perception of style or syuzhet information. His argument is not extended to the construction of the fabula, or the narrative of the film. I shall discuss the fabula in the

---

[45] Ibid. p.86.

[46] Ibid. p.31.

fourth chapter of my dissertation, but before turning to such high-order cognitive activities I would like to support and expand this direct or ecological view of the perception of images further.

Both Noël Carroll and Gregory Currie discuss our capacity to recognize images as an innate capacity, one which is naturally generative rather than conventional. Carroll cites the psychological work of Hochberg and Brooks, which suggests that as soon as children learn to recognize an object, they are capable of recognizing pictorial representations of that object.[48] Currie too, claims that in certain respects the experience of looking at a picture is like that of looking at the subject of that picture. Indeed, if one is presented with a picture of a horse

> To see that the picture is a picture of a horse, I deploy my horse recognition capacities. That is, I use the same capacity to recognize the picture of a horse that I use to recognize a horse.[49]

Recognizing the content of depictions thus involves recognizing the spatial features of the object depicted. Aside from realizing that a two-dimensional surface stands for a three-dimensional object, recognizing pictures does not involve learning. Since the spectator recognizes simply by looking, cinematic images comprise a very special kind of symbols. Carroll claims:

---

[47] Ibid. P.17.

[48] Hochberg, J. & V. Brooks, "The Perception of Motion Pictures," in Handbook of Perception, eds. E.C. Carterette and M. Friedman, vol.10, New York: Academic Press, 1978.

[49] Gregory Currie, op. cit. P. 80.

> The rapid development of this picture recognition capacity contrasts strongly with the acquisition of symbol system such as language. Upon mastering a couple of words, the child is nowhere near mastering the entire language. Similarly, when an adult is exposed to one or two representational *pictures* in an alien style, say a Westerner confronting a Japanese image in the floating-eye style, she will be able to identify the referent of any picture in the former after studying one or two representations of that sort for a few moments. But no westerner, upon learning one or two linguistic symbols of the Japanese language, could go on to identify the reference of all, or even merely a few more, Japanese words.[50]

As discussed in the last chapter, because of the survival role of the visual system, seeing has evolved as primarily a perceptual bottom up process designed to provide us with inferences about our environment as fast as possible. Biederman showed us that objects are decomposed into basic geons, and that this process precedes naming and categorization.[51] And Fodor claimed that because perception is encapsulated, it is sometimes dumb or insensitive to (top-down) cognitive knowledge.[52] A shot of a dinosaur in a film, thus, may be initially classified as a dinosaur by the quick and dirty perceptual mechanisms. Only once processed in the slower, deeper, global and higher cognitive process will it be understood as a depiction of a dinosaur, rather than the thing itself, since belief systems tell us that dinosaurs do not exist anymore. Similarly, a shot of a dead man will be understood as the image of a dead man, even though one knows the scene has been fabricated and that it is unlikely the actor has actually died during the shooting. Unlike language, the recognition of a realist pictorial representation of an object does not require a prior familiarity with the object. The ability to recognize objects may be acquired pictorially, rather than referentially. Currie claims that "my capacity to recognize depictions of *F*s and my capacity to recognize *F*s are one and the same capacity, however acquired."[53] This account explains our tendency to "see in" pictorial representations rather than "see as." We first and foremost see and recognize the object that is presented, and only secondly, and after cognitive deliberations, do we categorize it as a representation.

---

[50] Noël Carroll, op. cit. Pp. 139-140.

[51] See second chapter: section on geons.

Moreover, since visual object recognition relies heavily on the biological and neurological systems, rather than on information processing mechanisms, it behaves more like a reflex, and does not require "reading" or decoding for its inferences. And Carroll points out that

> [. . .] if the recognition of movie images is more analogous to a reflex process than it is to a process like reading, then following a movie may turn out to be less taxing, less a matter of active effort than reading.[54]

---

[52] See second chapter: section on bottom-up & top-down perception.

[53] Gregory Currie, op. cit. P. 86.

In addition, because realist movies are representations that rely for their deciphering on our biological wiring, rather than on learned practices, they become more widely accessible to most humans.[55]

One can see now that once the discussion of cinema is freed from the linguistic constraints of a semiotic study, and once cognitive evidence is accumulated, we can have a new and grounded understanding of how film perception actually operates. Following Anderson, then, the more film is constructed in accordance with our biological perceptual mechanisms, the more realistic its impression,[56] and the more easily it is followed by audiences. The cinematic image is not perceived as an iconic sign, a signifier which refers to its signified via a convention. It is perceived as a pictorial depiction of the object itself, one that is governed by the same perceptual biological mechanisms of seeing the object, rather than by a cultural convention about the object. While image recognition is fast and operates mainly at the perceptual low-level mechanisms, language comprehension requires high-order cognitive operations at work. As I showed in the last chapter, images are processed holistically and automatically, but also serially. In comparison, language is delivered and processed serially, and propositionally. While watching a film both auditory and visual sensors are being triggered. The perception of

---

[54] Noël Carroll, op. cit. Pp. 143-144.

[55] Ibid. p. 142.

[56] Both Currie and Carroll discuss the image while attempting to address questions regarding Realism as a cinematic style. While Carroll is interested in the non ideological base of the image, Currie is interested in deciphering the kind of illusion cinema is. Both discussions are fascinating, but outside the bounds of this work. It is important though,

images, and their further cognitive processing is quite different than that of language, whether delivered in the dialogue, as a voice over, or as verbal information in the visual track (written text, numbers, signs, etc.). This perception is further cognitively processed in cross modal operations, and is stored in memory towards the construction of the fabula. At the end of the last chapter I mentioned that cognitive science shows that images are stored in memory both holistically and propositionally. In the next chapter I will talk about how the memory of visual and verbal information affects the construction of the fabula. But before doing that, I would like to talk about film narration, or how images and sounds are organized in order to form the syuzhet, on which basis the fabula will be created.

Narration

In the first chapter I discussed film narratology, particularly in the context of literary narratology, on which it relies, and the latter's effects on the terminology and analysis conducted in film narratology. I followed Gerald Prince and others in treating the narrative as a product of both the story and its narration, or as the recounting of events by at least one narrator. The narrative also relies on the activity of the perceiver, and requires active participation of its audience in order to exist. The construction of the narrative and other perceiver related activities will be discussed in the next chapter. Here I would like to examine the narration

---

as Anderson pointed out, to show the connection between cinematic Realism and the ease of filmic perception.

of film, or how a film "tells" or recounts the events it portrays. Once we understand how the filmic story is being told (with lights and sounds), it will be easier to examine what the perceiver can and cannot do with that material presentation.

Gerald Prince defines narration as "the production of the narrative; the recounting of a series of situations and events."[57] This (literary) distinction alludes to an agent, a source, someone who is recounting the events. That agency can be the implied author -- the implicit image of an author in the text, the one responsible for the ideological or moral positions represented in the text.[58] But often, the agency behind the narration is regarded as the narrator "the linguistic subject, a function and not a person, which expresses itself in the language that constitutes the text."[59] Whether implied author or narrator,[60] the literary notion of narration involves an agency and a verbal utterance. Strict theorists like Gennette claim that narrative is a verbal transmission, an utterance, and therefore film, comic strip and the *roman photo* are "transmitted by an extranarrative medium."[61] Under this narrow view mimetic presentation

---

[57] Gerald Prince, Dictionary of Narratology, Lincoln and London: University of Nebraska Press, 1987, p.57.

[58] For a fuller account of the implied author see Wayne Booth, The Rhetoric of Fiction, Chicago: University of Chicago press, 1961.

[59] Mieke Bal, Narratology: Introduction to the Theory of Narrative, Toronto, Buffalo, London: University of Toronto Press, 1985, p.119.

[60] I realize, of course, that the discussion of the implied author and narrator is much more complex than what I portray here. As far as this discussion is relevant to film I will refer to it in greater length shortly. For a fuller discussion of these terms in literary narratology please see Wayne Booth, Rimmon-Kenan, and Seymour Chatman, in books quoted in this dissertation.

(i.e., drama – where information is delivered via dialogue) will be considered impure or mixed narration. Only diégésis, which is narration without dialogue, is pure narration.[62] In contrast, Rimmon- Kenan and others claim that the Platonic account includes both the mimetic and the diegetic as forms of narration.[63] This literary discussion is important since film is a unique form of a narrative. While it certainly qualifies as a narrative (a recounting of a set of events), it does not necessarily employ a narrator in the diegetic or the mimetic terms. Rimmon-Kenan defines narration as an event where information is transmitted about story events.[64] This definition, which stays away from both agency and the commitment to the verbal medium opens the door for a description of film narration. David Bordwell defines film narration

> As a process, the activity of selecting, arranging and rendering story material in order to achieve specific time- bound effects on a perceiver.[65]

Rather than focusing on the verbal and utterance aspects of narration, Bordwell (following Rimmon-Kenan and others) is interested in the discursive aspects of narration. Film communicates through images, dialogue, music and sound effects. Bordwell calls these channels of

---

[61] Gerard Genette, Narrative Discourse Revisited, trans. By Jane E. Lewin, Ithaca, NY: Cornell University Press, 1988, p.16.

[62] Ibid. p.18.

[63] Shlomith Rimmon-Kenan, Narrative Fiction: Contemporary Poetics, London and New York: Routledge, 1983, p. 106.

[64] Ibid. p.89.

[65] David Bordwell, Narration in the Fiction Film, Madison: University of Wisconsin Press, 1985, p.xi.

communication the film's style. Film narration is how the film's style is organized to create a syuzhet.[66] While interpreting a film, the audience is presented with a wealth of information, which is somehow organized (in the mind of the perceiver) to present the film's narration. Narration here, rather than being a comprehensive discursive linear practice from narrator to narratee and reader (*à la* Genette), is a product of an interpretive process of the perceiver. In his Making Meaning Bordwell attempts to explain what leads the audience to organize the information in a coherent manner. The leading factor is the interest of the perceiver of a drama in identifying whose story is presented, "in other words, it usually is of first importance to see whose futures are at stake – whose situation is settled by the events that are described."[67] In a section entitled "Making Films Personal" Bordwell makes a strong case to view films as a sub-set of large experiential cases whereby we personify objects in our everyday world in order to understand them better.[68] In interpreting films we apply our folk psychological intuitions about persons to filmmakers' intentions, and to characters. Bordwell calls this process 'the application of the person schemata,' which when activated is based on the following tenets:

1. A human body, presumed to be singular and unified.
2. Perceptual activity, including self-awareness.
3. Thoughts, including beliefs.
4. Feelings or emotions.

---

[66] Ibid. pp.58-59.

[67] David Bordwell, Making Meaning: Inference and Rhetoric in the Interpretation of Cinema, Harvard University Press, 1989, p.170.

5. Traits, or persisting dispositional qualities.
6. The capacity for self impelled actions, such as communication, goal -- formation and – achievement, and so on.[69]

The application of the person schemata is particularly important when attributed to the characters in the film. As I have shown in the first chapter, perceivers of dramatic texts tend to equate the characters in those texts with autonomous and subjective human beings.[70] And Bordwell contends that this dramatic tendency, in accordance with our general psychological tendency to personify our world, results in the character being at the focus of the narratorial process. Bordwell's idea of narration as a practice is based on the Bull's Eye Schema. At the core of the schema, the focal point of the interpretive process are the characters, with all the person attributes recounted above. A second circle describes the diegetic world, the surroundings of the characters and their activities. A third concentric circle represents the nondiegetic information like camerawork, editing, and music.[71] In defending the claim that the focal point of the schema is the characters Bordwell writes:

> [The schema] promotes those personified agents we call characters (fictional or not) over less prominent cues. The schema also suggests fruitful correlations: between character and setting,

---

[68] Ibid. pp. 151-168.

[69] Ibid. p.152.

[70] See chapter 1, pp. 15-16.

[71] David Bordwell, Making Meaning, op. cit. P. 171.

> between setting and camerawork. It thus offers a way for the critic to map semantic fields onto stylistic or narratorial qualities.[72]

In the following pages Bordwell gives countless examples of critical writings on films which regularly use this schemata to describe framing or lighting as motivated by a character's state of mind. Narration in this account becomes a selective process of the interpreter, first applied to elements related to the character, then to the diegetic world (which according to Bordwell is harder to notice), and finally to the extra diegetic properties of the film. As mentioned at the end of the second chapter, cognitive evidence can explain the trajectory Bordwell chooses. The activity of the characters -- both verbal, through the dialogue, and in visual actions -- is linear, causal and is therefore perceived and remembered in propositional or symbolic chains. But large spaces like rooms, sound effects emanating from those spaces, or framing and editing, are perceived quite differently. They are often perceived holistically, in a fast, crude and automatic process. Indeed, if a film "wants us" to notice an object in the frame (say a gun, for example), the object will either be foregrounded, well lit, or otherwise visually emphasized in the wide shot, or it will be granted its own close-up shot. Much as in real life, perceivers are presented with an overwhelming amount of visual data at once, and are unable to decode them all serially at once. We employ scripts (expecting certain items to be present in bedrooms and not kitchens) and other visual shortcuts to facilitate perception in the time provided. Visual pictorial information, as I have shown, is not processed as signs, not decoded as language, but processed much like the real world is, that is both holistically and serially. But we are much more aware of symbolic representations, such as language or linear action. Verbal perception, which is based on this highly coded information, is decoded, and processed serially, and then stored in memory propositionally. In this form, it is much more available for high-order cognitive activities such as causal ordering, reasoning, and other mental operations. Verbal information is therefore more readily useful than information derived from images when constructing the narrative. Similarly, in the visual field, characters' actions are easily reduced to causal and linear chains, therefore lending themselves to a serial interpretation, and to propositional encoding sets, for memory storage purposes. Bordwell's account, like the semioticians', prioritizes the (high-order) coded channels of information, over the (low order) holistically perceived ones.

---

[72] Ibid. p.170.

While Bordwell makes a strong case for evaluating the narrative from the core out (i.e., from character out to extra-diegetic worlds), he is willing to see that in cases like Godard's cinema -- where characters are alienated from audiences by design -- this direction may not work. He then suggests

> The simplest textual heuristic, then, seeks synchronic correspondences between the cues in different rings of the circle. A more dynamic heuristic treats meaning as having its primary source in either the innermost circle or the most enveloping one,

> and as passing "through" the others. In other words meaning "moves" outward or inward.[73]

Bordwell's account is interesting and useful, but is puzzling in two related areas. The schema is designed to give coherence both to the subjectivity of the character, and to that of the narrative. Bordwell's model of narration is aimed at unifying the stylistic elements (both diegetic and extradiegetic) in support of a unified subjectivity for the character. That is, the Bull's Eye schema, if applied properly, reiterates the person schema, thus stabilizing the meaning of the text in its characters. While this tendency can be written off as a somewhat romantic and nostalgic notion when applied to the characters, its subsequent result, the unification of the overall narrative, is more bothersome. Bordwell's model of narration only picks up details that support the person schema, or that cohere according to some other enveloping principle (such as a Godardian aesthetic ideology). But this must be a reductive and simplistic view of narration, especially if it is supposed to apply to all narratives. A film can tell us a story based on a set of contradictory narratives, such as the case in *Rashomon* (Kurossawa, 1950). Or a film can set its audience up with an inscribed subversive narrative (or a few), a tactic common to crime and detective films. How can the Bull's Eye Schema explain the narration of films that are not so easily unified? According to Bordwell's account there will always be some cohering narratorial principal. And while one can find philosophical or ideological explanations to *Rashomon* (such as existentialism as the reason for the open-ended narrative, for example), it is hard to reconcile the film as a coherent narration. And how will we understand the ending of many postmodernist films such as *Breaking The Waves* (Lars Von Trier, 1997), and *The Usual Suspects* (Brian Singer, 1995)? Those films are aimed (I would even dare say narrated) so as to create competing alternative readings where a sense of closure, coherence, and unity are not only evaded, but loudly denied. While Bordwell cautions us that "to give every film a narrator or an implied author is to indulge in an anthropomorphic fiction,"[74] and while he insists that if a narrator is constructed it is "the product of specific organizational principles, historical factors, and viewers' mental sets,"[75] this interpretive construct is always referred to as a unified, singular entity. And such a unified function cannot account for the ambiguity mentioned above. Later in this chapter I shall propose a model of narratorial ambiguity, one that will account for the multiplicity of cognitive processes at work in narrating and interpreting a film. But before doing so, I would like to review a more complex concept of narration, that of Edward Branigan.

---

[73] Ibid. p.181.

While Bordwell refers to narration as a discursive act, as means of interaction between style, syuzhet and the fabula, Branigan approaches narration epistemically. While the narrative contains the "what" of the story, or what Branigan calls the declarative knowledge, "narration addresses issues of procedure: *how* are we acquiring knowledge about what is happening in the story"[76] or whether there are conflicts in the procedures through which information is provided. Compared to Bordwell's somewhat psychological schema, this account is structural at heart: it refers to levels of narration, which may co-exist, rule each other out, or be in tension with each other. Narration is based on the principle of uneven distribution of knowledge, whereby one character knows more than other characters, and may know more or less than the audience. Branigan writes:

---

[74] David Bordwell, Narration in the Fiction Film, op. cit. P.62.

[75] Ibid.

> [. . .] narration will be comprised of three elements: a *subject* in an *asymmetrical* relationship with an *object*. As we shall see, the perceiving "subject" may be a character, narrator, author, the spectator, or some other entity depending on the context that is being analyzed.[77]

Moreover, the subject at one time in the film, may become the object of another epistemic setup. For example, in *The Silence of the Lambs,* Jack Crawford (Clarice's FBI superior) is at one moment informing her and the viewers that he is on the way to the serial killer's home in Michigan. He is the subject, and she is the (disappointed to be left out of the action) object. But only minutes later we realize with her that she is at the killer's house, thus when Crawford breaks in to the empty house in Michigan, the audience is the subject, and the FBI officer the object in an epistemic exchange (see appendix #5).

---

[76] Edward Branigan, Narrative Comprehension and Film, New York and London: Routledge, 1992, p. 65.

[77] Ibid. p.66.

Branigan's account enables us to have (a) complex, or even contradicting narratives, and (b) a concept of narration that is not all encompassing or unified, but is at any given moment susceptible to unreliability, or to change in the epistemic positioning. Branigan defines eight levels of narration, starting with the historical author, the extra-fictional narrator, the nondiegetic narrator, diegetic narrator, character (nonfocalized narration), external focalizer, internal focalization (surface), and internal focalization (depth).[78] These narrators are matched by a series of narratees, both fictional and implied audience, which will be discussed in greater length in the next chapter. The levels of narration also correspond to levels of representation in the artistic text, from historical information to the internal dynamics of a character's mental world. Those textual levels are the text, fiction, story world, event/scene, action, perception and thought. The perception of information by the viewer is always subject to questions of what is being communicated and under which conditions. The text becomes a composition of "a hierarchical series of levels of narration, each defining an epistemological context within which to describe data."[79] Branigan devotes much of his book to an analysis of film narration based on this model. He incorporates his earlier work on point of view[80] into a discussion of subjectivity, point of view, and other narratorial concerns.

---

[78] Ibid. P.87.

[79] Ibid.

[80] See Edward Branigan, Point of View in the Cinema: A Theory of Narration and Subjectivity in Classical Film, Berlin, New York, Amsterdam: Mouton Publishers, 1984.

Branigan's model is very comprehensive and well thought out. Our narratorial judgement and cognitive processing are a product of our estimation of the interplay between the epistemic levels of narration. But Branigan refrains from distinguishing how the information is provided. Whether information is available through action or speech, it is evaluated based on its narratorial source –internal or external focalizer, or just a non-focalized character – and not based on its cognitive processing. Yet, as I have shown in the previous chapter, the channel of communication (verbal or visual) has impact on its cognitive processing. Visual perception is processed differently from verbal perception, is treated cognitively differently, and remembered somewhat differently. While there are many cross-modal operations at high-level cognitive activity such as the construction of the narrative, some early perceptual judgements (e.g., what to focus our visual attention on) may have substantial implications for narrative construction skills. If we haven't noticed the gun on the counter, we may be unable to predict all possible scenarios for the future action of the characters involved. In other words, while Branigan's notion of procedural knowledge is concentrated on the structural aspects of the epistemic boundaries (who is telling, and how do they know what they are telling), it seems that cognitive evidence about perception suggests that we should also take into account style, or how the information is being presented. Whether character or perceiver, the source of the information (i.e., verbal or visual) is crucial for our understanding and judging the importance of that information. Branigan's concentration on the syuzhet, while ignoring style (to use

Bordwell's terminology), is a detriment of his theory. After all, it is very hard to talk about the syuzhet as separate from style, given that style is the material presentation of syuzhet information. Let us consider an example.

One of the cinematic devices to convey the attribution of knowledge to characters is point-of-view editing. Branigan describes the point-of view structure as a set of two shots: the first, the point/glance shot, reveals a character looking (usually off camera), and the second, the point/object shot, assumes the physical point of view of the character, or simulates it, to show the object as seen by the subject.[81] In the first chapter I claimed that a third shot is often necessary to complete the effect, a reaction shot (point/glance 2) in which we see the character's response to the object/character/event seen. An example can be seen in *The Silence of the Lambs*, in the scene where Clarice has just entered the house of the serial killer. The implied author has shown the audience in the previous scene the starving girl in the well, and her success in seizing the killer's dog. When the doorbell rings and the distressed killer opens it to an unsuspecting Clarice, the audience is in a position of epistemic advantage over her. We know she is at the killer's and is therefore in danger, but she does not know that. Clarice walks into the kitchen, her gaze wanders around the cluttered space, finally settling on something. The camera shows a close-up of a butterfly, therefore hinting at a detail known about the killer – that he uses rare butterflies from South America. But the audience already knows that Clarice is at the killer's, so the important narratorial question is: does Clarice know? The point/glance, point/object pair of shots is then followed by a third shot, that of Clarice with her mouth slightly open in surprise, and her hand, slowly and quietly moving to rest on her gun.

---

[81] Ibid. p. 103.

I shall return to the issue of the number of shots in a point of view set while analyzing *Rambling Rose,* but for the moment, I would like to concentrate on the theoretical formulation of the point of view device according to Branigan. Branigan writes:

> The theory that point of view is an attitude of some sort – of a viewer, implied observer, character, narrator, implied or real author, etc. – is based on the assumption that the bearer of the "attitude" is "like," or simply is, a real person who expresses himself/herself in a way which communicates with the viewer as a person. By contrast, I will be dealing not with e.g., characters and their (?) attitudes but with codes of character and the ideology of the text. The concept of psychological attitude is wholly replaced by those of textual production and ideology. I will take point of view to be a property of a language system, thus working toward a general theory of representation in film.[82]

Since Branigan's agenda is to look at point of view as a structural device in the creation of the narrative, it is clear why he attacks romantic notions of character subjectivity. But the downside of tending so heavily toward a structural explanation is that it ignores some basic cognitive tenets active during viewing. I have already mentioned Bordwell's (or folk psychology's) person schemata, in which we personify objects and aspects of our life in order to more easily process them. Moreover, psychological research shows that infants of two to three months of age start following their mother's gaze stabilizing it onto objects the mother is looking at.[83] This data is easily explained, as following the glance of another human being or a predator could reveal the intentions of those subjects. That is, the tendency to follow the gaze is biologically ingrained in human beings as one of the survival mechanisms. Perceivers are sometimes presented in film with the wandering gaze of a character, then the camera cuts to a pan or a tilt movement through space, which is attributed to the scanning gaze of the character. But in point of view editing, the actual camera movement between gaze and target is usually deleted, and the audience "fills in the gap" assuming the character was actually looking and the camera just picks the image up once the gaze was settled on an object. Point of view editing not only reveals intentions, but is also instrumental in revealing emotions. Recall the opening scene of *The Elephant Man*, in which Dr. Fredrick Treves (played by Anthony Hopkins) sees the Elephant Man for the first time. The Elephant Man is slowly turning to face the doctor, and he is obscured by a cape, and dimmed lighting. The film cuts away from the turning figure before he arrives to the full frontal position, so the audience is never granted a view of the Elephant Man (in fact, not until the very end of the film). But the film shows the face of Hopkins, the camera zooming in on it, as a tear forms in his eye. The horror of the sight, and the empathy the doctor feels are transmitted to the audience without ever seeing the subject of the film. Identifying the emotions of the characters and their target objects/causes is of primary importance to the understanding of the narrative. Noël Carroll claims:

---

[82] Ibid. p. 17.

---

[83] George Butterworth and Lesley Grover, "The Origins of Rreferential Communication in Human Infancy" in (ed.) Lawrence Weiskrantz, Thought Without Language, Oxford:

> Stated boldly, point of view editing can function communicatively because it is a representational elaboration of a natural information gathering behaviour. That is, point of view editing, of the prospective variety at least, works because it relies on depicting biologically innate information gathering procedures. This is why the device is so quickly assimilated and applied by masses of untutored spectators.[84]

We can see now that while it is possible to support Branigan with his attempt to not attribute subjectivity to textual entities such as characters and narrators, it is important to acknowledge the cognitive mechanisms at work in the process of actual viewing. Treating point of view as a structural device (in a language-like system) is cumbersome, requires many complex cognitive mechanisms, and cannot explain the ease with which this device is used and processed. While for Branigan point of view merely "describes the correlation between two narrations"[85] a cognitive account can explain psychological mechanisms of identification and

---

Clarendon Press, 1988, pp. 5-24.

[84] Noël Carroll, Theorizing the Moving Image, New York, NY: Cambridge University Press, 1996, p. 129.

interpretation which are rooted biologically in our survival mechanisms. In other words, in accordance with Occam's razor, the cognitive account operates at lower perceptual levels, is simpler cognitively, and should therefore be preferred theoretically. I do not claim that Branigan's theory of narration is wrong, only that it is partial. While his structural model is very useful, one also needs to examine style, and its cognitive effects on syuzhet, which in turn will affect the construction of the fabula. I do believe though that the solution to the limitations of Branigan's theory can be found in a logical extension of his own model.

Branigan's notion of narration is based on embedding: one level's narrator becomes the object of another. This model allows for several levels of narration to operate simultaneously, so as to encourage compatible interpretations. Moreover, this model, which is inherently hierarchical, does not end with ultimate epistemic saturation. For Branigan:

> Omniscience does not mean that the reader finally knows all, or that there is an author/narrator who knows all, but merely refers to the reader's toleration of a boundary, or *limit* to what finally can be known *in* the text.[86]

Branigan's model is much more open to ambiguity, unreliability, or compatible interpretations than most literary models. Unreliable narration usually refers to a narrator whose values and claims are

---

[85] Edward Branigan, Point of View in the Cinema, op. cit. P. 177.

[86] Edward Branigan, Narrative Comprehension and Film, op. cit. P.115.

undermined by the implied author. In this case the perceivers receive information that is then deemed to be untrue, or probably not true, but there is no way to determine what really happens. This approach assumes that the unreliable narrator's discourse is embedded within another level, of another narrator, who eventually unifies the text by proving to the readers that the unreliable narrator was such. But we have seen throughout the first chapter and this one that narration in film is often done without a narrator, at least in the strict sense of the term. Gregory Currie writes:

> But I shall argue that there are narratives which are unreliable even when there is no narrator. In these cases, unreliability is not the product of a disparity between two conflicting viewpoints, one internal (the narrator's) and one external. Rather it is the product of a single, external viewpoint which has, as we shall see, a rather complex structure.[87]

For Currie, this external perspective is provided by the implied author (sometimes also called the external narrator). While film has embedded narrators (like the people telling their stories about Charles Kane in *Citizen Kane*), it is difficult to import to film the literary idea of a controlling narrator. A controlling narrator is some fictional entity who is assumed to relay all of the story's events to the readers, but is not the implied author (or filmmaker). But it is hard to imagine what form such an entity would take in film. Even in the cases where we have voice over narration, the film is not just told verbally, it is shown, heard, played musically, in action and setting, and so on. Very few narrative films show the entire events without showing the main character (Tarkovsky's *Mirror* is the only one that comes to my mind). And as soon as we see the main character, someone is narrating the film. But this narrating agency is not a unified subjectivity. In literary fiction, one may collapse the author, implied author, and controlling narrator to one entity. But it is hard to collapse those roles in film. The audience is aware of the numerous agencies (cinematographer, editor, set designer, etc.) behind the "telling" of the filmic story. It seems implausible therefore that audiences would attribute such a literary controlling role to any one diegetic filmic entity.[88] And if we abolish the role of the controlling narrator, filmic unreliability can only be the product of embedded narrators. Yet Currie gives examples of films in which narration is unreliable, but the source of the unreliability is not the embedded narrator.[89] In such cases Currie claims that there is a complex intention on the part of the implied author, and he calls that narration ambiguous. While unreliability is usually negated at narrative closure -- that is, the text's fabula is unified -- ambiguity is much more open. A narration is ambiguous "when it raises a question in the reader's/viewer's mind which it fails to answer, and where the raising and the nonanswering seem to have been intentional."[90] In ambiguous narration, the text lends itself to several interpretations, and it refrains from giving the perceiver enough information to decide which interpretation is preferred. Complex unreliability does not necessarily have to be ambiguous, but the latter is easier to achieve. To set up complex unreliability the implied author needs to plant clues at two levels

---

[87] Gregory Currie, op. cit. p. 261.

---

[88] Ibid. p. 267.

[89] Ibid. pp. 269-280. Mostly an analysis of Fritz Lang's *You Only Live Once*.

> [. . .] at level 1, where the clues are more obvious, but only superficially persuasive; and at level 2, where they are less obvious but more weighty when reflected upon.[91]

An unreliable narration implies a transition from openness in the first level to closure on the second: the narrator has been identified as unreliable. And deeming the narrator unreliable is usually done by another level of explicit narration. But in film, given that there is no controlling narrator, and most narration is done by an implied author, the task of closure is much more difficult than a move toward opening, or ambiguating the situation at the second level. Currie never explains why it is easier for unreliable narration to exist in literature and for ambiguous narration to exist in cinema. But I would claim that a cognitive approach can shed some light on the matter. Sol Worth points out that unlike words, pictures cannot negate, all that pictures can show is what *is* – on the picture's surface.[92] And in light of cognitive data this is quite clear. Pictures, as I have shown in this chapter, are perceived by the same biological mechanisms used for object recognition. Currie calls this the proposition of perceptual realism, by which pictures are seen like the

---

[90] Ibid. p. 274.

[91] Ibid. p. 276.

things they represent. An image of X then is seen as an X. For an image to be able to negate itself, there would have to be a representational distance between the object and the picture, and there would have to be a convention that enables the image to postulate the claim "is not." But those properties, as I have shown before, belong primarily to natural languages (somewhat also to coded visual signs, like road signs), but do not have equivalencies in visual photographic images. For visual film narration to negate, it will have to employ another mode of narration like dialogue, or the use of editing conventions. In *The Silence of the Lambs* for instance, Jonathan Demme uses both methods to mark the implied author as unreliable. The scene in which Clarice is in Ohio and the FBI are heading to the serial killer's house in Michigan is organized so as to lead the audience to believe that Clarice is out of the "action" (see appendix #5). First, we hear verbally that the FBI has the killer's address and are presently flying out there. Next, we see a house, and a teleprompter title comes on the screen, naming the city in Michigan (shot #1). The camera cuts to the inside of the house where a series of shots conveys that the serial killer is playing with bugs (shots 2,3, & 5). A series of cross cutting between the exterior of the house where the FBI agents are getting ready to break in (shots 4, 8, 12, 15, 18, 23, 25, 29) and the interior where the serial killer panics when he realizes that the kidnapped girl seized his dog (7, 9-11, 13, 14, 16, 17, 19-22, 24, 26-28), follows. This alternating syntagma (which in film convention designates different spaces, same time)

---

[92] Sol Worth, Studying Visual Communication, Philadelphia: University of Pennsylvania Press, 1981, p. 174.

culminates in the typical spatial unification of inside and outside when the FBI agent rings the doorbell twice (shots 30, 35), the camera cuts to show the doorbell from inside (shots 31, 36), and the serial killer responds. But the implied author was leading us down the wrong path and when the serial killer opens the door and (in an over-his-shoulder shot #39) we see Clarice, we immediately realize we have been fooled. We then proceed to "correct" the logical reasoning, adjust our frame of reference, and proceed with the film's events. Unreliability here was clearly exercised, but given that it was performed by the highest authority of the film, its implied author, the audience cannot judge that narration, but can only accept they have been fooled and move on. Moreover, this unreliability was achieved not visually, but by conventional visual arrangement (parallel editing, or alternate syntagma) and by the verbal messages printed on the screen, thus contextualizing the images in a (wrong) geography that is crucial to the narrative's punch line. But this unreliability does not seem to destroy the authority of the implied author. As soon as we have performed the logical adjustment (see section on belief revision in chapter 4), we accept that a trick was played on us, and we return to judging the following information as reliable. The reason, again, is that in realist cinema, as audiences we have no way of deeming the visual information as unreliable, unless it is contextualized by an extra visual medium as such. For that very reason, the conclusion we draw at the end of *The Usual Suspects* is quite the opposite. The film is comprised of a visualization in flash back, of a story a character tells while being interrogated at the police. When we learn at the end of the film that the character fabricated the story using names and facts which were only available to him only at the investigator's office, we deem him as unreliable internal narrator. The visual story though, was only an illustration of a verbal one, and verbal accounts can very easily lie. But the film does not end there, as the suspect who has just left the police station is now being picked up by one of the (supposedly dead) main characters from his story. The credits start rolling as the car pulls away, and the police officer arrives huffing and puffing at the street corner. The film then leaves the epistemic status of what we have just seen unresolved. The implied author never reveals whether it was a fabrication, or truth, or which parts of the story were true, that is, true in the diegetic fictional world. Moreover, at the end of the film the audience is not sure whether the film's implied author was unreliable, or whether it is just unreliable embedded narration. The narration then is at best ambiguous.

Ambiguity, unlike unreliability, is much easier to achieve visually. Moreover, the visual track and the dialogue track may propose alternative narrations, ones the implied author may choose not to embed and hierarchize. In *Eyes Wide Shut* (Stanley Kubrick, 1999) Bill (played by Tom Cruise) and Alice (played by Nicole Kidman) are engaged in a heavy argument. They are arguing about fidelity, deception and their relationship. Bill is guarded, uses stereotypes, and lies. He boldly tells Alice he trusts her not to cheat on him. Alice responds with a story about a vacation they took together a year earlier. The story is about a sexual fantasy, one she has never acted upon. She seems honest, and in the narrative logic of the argument, she has no reason to lie about it. But her body language and the tone of her voice tease Bill to believe she may have acted out that fantasy. For the rest of the film Bill is tormented with images which enact the fantasy. Because the film deals with masks and masking in intimate relationships, Kubrick never identifies whether Alice was honest or not. The two narratives, the one of body language and the one of verbal information, are never presented in a larger context, which embeds and hierarchizes those narrations, indicating that one is true, and the other false. Moreover, the visual enactment is clearly marked as Bill's fantasy/nightmare, that is, its epistemic status is quite weak. Kubrick wants to leave the film openly ambiguous, hence Alice's last words telling Bill they need to fuck, and the film fading to black without showing the audience Bill's response. Looking at the tracks of communication, or style, and how they are perceived cognitively can explain ambiguity, which seems to be more and more prevalent in film. The logical extension of Branigan's theory would be to treat procedural knowledge not just as it pertains to the content, but also as to how this knowledge was achieved, that is to ask not only who tells us what, but also was it told visually or verbally and what are the implications of each.

A cognitive approach to point of view

In his insightful structural analysis of point of view, Boris Uspensky provides a very useful tool.[93] Uspensky discusses point of view not merely as the physical view point of a character, but as a structural device that follows the emotional, psychological or ideological state of mind of a character. Thus, *The Conversation* (Coppola, 1974) subordinates the entire narrative to Harry Caul's mental point of view. At the end of the film the audience finds together with Caul, that his understanding of the state of affairs was wrong. At no point in the film does the audience know more than Caul, that is, the film's narration is subordinated to an embedded narrator's epistemic constraints. Moreover, even when the film shows Caul's hallucinations/ nightmares, it refrains from asserting them as such. That is, according to Branigan, the film is told from the deepest level of narration, the internal focalizer.[94] But according to Branigan, that level should lead to identification with the character/narrator, and that is not the case in *The Conversation.* Caul's profession is surveillance, but the film rarely shows things from Caul's physical view point, or perspective. In fact, the film rarely shows Caul in a close-up, or even a medium shot. Most of the film is shot with a wide-angle lens, therefore making it difficult to determine how Caul really feels. Given that Caul is very protective of his privacy, this visual choice makes narrative sense. But as a result, Caul as a focalizer is weak at least as far as identification with him takes place. And while epistemically the audience is relying on Caul as the sole source of narrative information, emotionally the film maintains a distance between audience and Caul, a distance that makes it hard to accept the film as one that exhibits a high degree of reliance on a point of view. While Uspensky's and Branigan's theories cannot explain this heightened distance between audience and character, a cognitive approach can. Physical point of view structure (comprised of a series of three shots: point/glance, point/object and point/glance) initiates – as I have shown earlier -- an innate biological procedure used to understand the world around us. This narratorial structural device appeals to our most basic, low-level perceptual practices. It reveals intentions and emotions, and enables us to see others' responses to a given situation. But understanding the mental point of view structure such as the one employed in *The Conversation* relies on verbal cues, and on a grasp of the overall narrative premise, both of which involve high-order cognitive mechanisms. I do not try to claim here that focalization is always more effective when it conveys the physical view point, rather than a mental one. A film like *Vertigo* (Hitchcock, 1958), successfully conveys mental point of view, but it is successful precisely because the device to portray this mental perspective is a visual point of view structure. That is, we understand Scottie's fear of heights once we see things, so to speak "through his eyes." Once the coding of his mental state was done visually, we do not need reminders (i.e., more point of view shots), but we can see things from his mental perspective. Moreover, traditional shots of the city from above, which would normally be seen as objective description of space, and as originating from the implied author level of telling, are not read anymore as either objective or as external narration, but as psychological point of view. Focalization here is predominantly done through a mental point of view, but that is anchored, at least once, in a physical view-point. Means of focalization are not the only challenge to Branigan's model of narration. Branigan claims:

---

[93] Boris Uspensky, A Poetics of Composition, trans. by Valentina Zavarin and Susan Wittig, Los Angeles, CA: University of California Press, 1973.

---

[94] Edward Branigan, op. cit. P.103.

> In a strict sense, a narrator offers statements *about*; an actor/agent acts *on* or is acted upon; and a focalizer has an experience *of*. More precisely, narration, action and focalization are three alternative modes of describing how knowledge may be stated or obtained.[95]

But in his model internal focalization evokes identification and external focalization (through both action and speech) evokes observation and character understanding. I have so far shown that statements (whether visual through editing, or verbal,) and actions, tend to require a serial and propositional processing, which are high-order cognitive operations. But the term focalization here is vague. Thoughts in cinema are conveyed either verbally by a synchronized or voice-over narration, or visually through body language and expressions. Experience of (or perception) seems to be indicating a more direct visual reception, although it may be conveyed through complex editorial devices such as in the case of *The Conversation*. Moreover, what may seem to be a shot focalized through a certain character can be also understood as an implied author's narration about another character. Branigan embraces the notion that "several narrations may be operating simultaneously"[96] but attributes the multiplicity to the different levels of narrations, and not to the source of the material.

---

[95] Ibid. p.105. Emphasis in original text.

[96] Ibid.

Rambling Rose

In the rest of the chapter I will suggest an extension of Branigan's model by performing an analysis of the modes of narrations in a few scenes of *Rambling Rose* (Martha Coolidge, 1991). The film follows the coming of age of Buddy (played by Lukas Haas), a 13-year old who falls in love with his nanny, Rose (played by Laura Dern), a sexually uncontrollable young woman. While Buddy is the main narrator, the film occasionally allows Rose to focalize but not narrate part of the story. More importantly, while Buddy controls the visual access of his sister (Doll) to the sexual action, the implied author occasionally controls Buddy's access too. The film uses point of view structure to control and restrict the characters' access to knowledge. The point of view structure also creates a complex set up of characters' subjectivities. A cognitive analysis of narration will reveal those mechanisms at work.

The film starts with exterior shots of a car driven by a man in his forties down the highway, superimposed with a title "Glenville, Georgia, 1971" and a voice over of the man. He explains that he is coming south to visit his father, and as he pulls in front of an old house, a wave of nostalgia is rushing over him. As he is stepping up to the porch the voice over says "When I was 13 a girl came to live in this house. She was the first person I loved outside my own family." The camera cuts to a point of view shot from the porch down, and in the distance we see a woman crossing a bridge which leads towards the house. When the camera cuts back to the porch we see Buddy, a 13-year old, looking at the woman. This opening sequence asserts the 40-year old Buddy as the narrator, and the 13-year old Buddy as the focalizer. Even though the film does not use voice over or show the 40-year old Buddy again until the very end, it is clear that this is Buddy's story. But the shots of the driving car, the title indicating the time, and the editorial manipulation into the flash-back assert another narratorial entity, the extra-fictional narrator (in Branigan's terms) or the implied author (in Currie's). The dynamics between this implied narrator, and Buddy as the focalizer/narrator is the topic of the following analysis. Back to the first scene: Buddy moves closer, and looks at a young woman in 1940s style dress walking up to the house. She doesn't see him at first, and as he gazes at her, he smiles. The camera cuts back and forth showing Rose walking, camera tilting-up scanning her body, and Buddy in a close-up looking without being seen. This traditionally voyeuristic structure privileges the male gaze, and Buddy's point of view shots emphasize that he sees Rose as a (sexual) woman. When Rose finally sees him, the shots mirror each other. Rose is walking left to right, camera tracking with her, Buddy moving right to left, camera following him as well. Rose is the first one to speak, introduces herself, and the scene ends before Buddy replies. The majority of the film from now on will show Rose as seen from the physical and emotional perspective of Buddy. Neither Rose, nor the rest of the family (with the possible exception of Doll) seem to be aware of the emotions Buddy develops towards Rose, which only increases the sense of voyeurism promoted by the film. The film's main narrator is indeed Buddy. But the story is Rose's story: a young woman, almost pushed into prostitution, who is saved by the Hillyers when invited to work at their house. She is nearly a nymphomaniac, so her sexual activity and the punishments she endures for it, are at the center of the plot. The second scene introduces Rose's first love object -- Mr. Hillyer (played by Robert Duval).

In the scene Rose is introduced to the family by Mrs. Hillyer, who tells her stories about each one of her three children. In the middle of the conversation Mr. Hillyer walks in to the house (see appendix #1). We hear him in voice over (placed alongside Rose's image, shot #1) and then see him briefly. The camera cuts back to Rose, a melodic flute tune fades in, and a camera zooms-in to close on Rose's face, as her gaze scans Mr. Hillyer's body (off frame, and to the right) and her mouth gapes in pleasant surprise (shot #3). Throughout the shot we hear Mr. Hillyer exclaim (#4): "well, well, well, so miss Rosebud has arrived." The camera cuts back to Mr. Hillyer (#5-6) who continues: "Rosebud, I swear to god you are as graceful as a capital letter S. You will give a glow and a shine to these old walls." The camera cuts to Rose as she blushes and smiles. Mr. Hillyer sits down and continues (#11):

> Now. It is my dear wife's belief, which I accept although I do not totally grasp it, that to hire a person to do household work is a. . . is a criminal practice. You are therefore here as a friend, as a guest, and indeed as a member of this family. In love and harmony, Rosebud, in love and harmony.

Rose's eyes fill with tears (#12, 14), and Mr. Hillyer continues to talk about Rose's past alluding to an attempt to draw Rose into prostitution in Birmingham, but assuring her that she is now in a safe haven. The sequence is edited by cutting back and forth between Rose and Mr. Hillyer (both in medium close-up shots). Only once does the camera cuts to show Buddy (#16), when the scoundrels in Alabama are mentioned. Mother is not seen at all, except for the introductory wide angle shot (#5), in which she is seated with her back to the camera.

Not so apparent, but very interestingly, all of the shots of Rose are shot from the physical view point not of the sitting Mr. Hillyer, but from that of the standing Buddy. Given that the film is set up as a coming of age drama, and the voice over of the adult Buddy already declared that Rose was the first person he fell in love with, we know that we physically see Rose as Buddy sees her: that is, through the eyes of an internal focalizer who experiences the events. Cognitively, mentally, and emotionally we are placed in Buddy's shoes, as the physical (traditional) point of view shot represents. But I would like to claim that the shot where the camera zooms in on Rose (#3) also functions quite differently, and at a different level of narration.

The camera, or actually the lens' zoom movement, is a change of the focal length of the lens, thus widening or narrowing the view seen, literally changing what the frame encompasses spatially without editing. It is a camera activity that like others (track, crane, pan, and tilt) draws attention to the technical aspects of filmmaking, that is, to the constructed nature of the film. It is usually used in accordance with a movement in the frame, so that the abrupt and mechanical change of frame size is not so apparent. But in the case of the shot described above there is no movement in the frame, so the camera movement calls attention to a (heavy handed) storytelling technique. As a result, aside from the internal focalizer's (character) narration, there is also a presence of an implied author. This external narrator is using the cinematic indirect discourse technique, as identified by Jeffrey Rush and mentioned in the first chapter of this dissertation.[97] Rush claims that physical point of view shots can be equated with direct discourse, or the quotation marks used in literary fiction. In those instances, the narrator seems to disappear as the character speaks her own words. In indirect discourse the narrator comments on the thoughts of the character, rather than imitating them.

> In indirect discourse, there is a tag, an attribution of thoughts to a character, but the narrator is responsible for summarizing those thoughts. That summary breaks down the boundary line of quotation marks, and brings the narrator and character closer together. Both are represented in the text.[98]

The zoom movement seems to suggest that an external narrator is describing Rose's response as if in third person. A verbal description of the shot will sound more or less like this: "Rose looked at Mr. Hillyer and thought to herself how handsome he was." In such a structure Rose becomes an object in the narrator's discourse. But the zoom movement in this shot does not function alone; it is accompanied by romantic extra-diegetic music, and by Rose's overt gaze which objectifies Mr. Hillyer. And if she is an object of the extra narrator's discourse (by zoom and music), she is also a subject capable of objectifying and desiring Mr.

---

[97] Jeffrey S. Rush, "Lyric Oneness: The Free Syntactical Indirect and the Boundary Between Narrative and Narration," in Wide Angle Vol.8 No. 3&4, 1986.

[98] Ibid. p. 29.

Hillyer. I would therefore like to claim that this shot can stand as representing Rose's mental/emotional point of view, as she falls in love with Mr. Hillyer. While we do not see things from Rose's view point, the music and her gaze indicate that she is a focalizer of sorts. Moreover, the zoom shot is edited between two shots of Mr. Hillyer. That is, an application of the point/glance, point/object, point/glance structure of point of view will show that the reverse happens, but is nearly as effective. What we have here is a point/object (Mr. Hillyer) followed by a mutation of point/glance shot (rather than Rose's physical view point we see her gazing up and down the object), and a return to point/object, making sure the audience knows what Rose was looking at. The reason that this mutation of the point of view structure is so effective is that this trajectory is similar to the trajectory of the baby's look at his mother's gaze and attention. As shown earlier, the cognitive understanding of the mother's focal attention precedes the understanding of point of view structure. Repeating that experience should therefore be as easily understood as placing the viewer in the physical view-point of a character. This leads to the conclusion that the sequence can be understood both as focalized through Rose and as focalized through Buddy.

In addition, I would like to suggest that there are three levels of narration at work: the implied author and Buddy both see Rose as an object, but Rose asserts her subjectivity through focalizing her own part of the story. She is not allowed to narrate it, as it is presented through indirect discourse, but her emotional point of view is asserted despite the double narration (of Buddy and the implied author). Branigan says:

> In general, several levels of narration will be operating simultaneously with varying degrees of explicitness and computability; that is, the spectator may describe the text in several different ways, all of which may be accurate, each within a particular context and for a particular purpose.[99]

While Branigan welcomes simultaneous narrations here, his model is hierarchical in nature, and if we follow the model and ask "how do we know what?" we should arrive at a clear distinction of levels of narration, or of the shifting subject/object relations between different levels of narration. But the scene described in *Rambling Rose* cannot be presented in such a hierarchy. While Rose's focalization is definitely an object to the (higher level) narration of the implied author, it exhibits no such relations of power to Buddy's narration. Rose's and Buddy's focalizations exist on the same epistemic level of the film, because they both are understood by a similar application of natural point of view procedures. Buddy's narration and the implied author's are similarly placed on the same plane. The implied author never sanctions or subordinates Buddy's perspective to her own. It will do so later in the film, when the content will be such that Buddy's initiation into proper "manhood" requires he does not know what is happening, and I shall return to that later. But in the meantime, we may say that Buddy and the implied author are busy telling two different stories.

---

[99] Edward Branigan, op. cit. P.96.

Some may claim that because Rose is only a focalizer and Buddy is both a focalizer and a narrator, his status on the hierarchical epistemic totem pole is higher. It is important here to understand that the film underplays the construction of Buddy's point of view. While a floor plan of the room reveals that Rose's images are shot from Buddy's perspective, the film does not provide a point/glance shot (i.e., Buddy's look) before or after Rose's, so it is hard to register the shot as Buddy's upon first viewing. The cutting back and forth between Rose and Mr. Hillyer narrows the focal attention to the two of them. Moreover, as I have shown in the second chapter, we use mental maps to grasp large scale environments. Given that Buddy is in the dark background from the beginning of the scene, and his placement with regards to his father is somewhat obscure -- at one point we even have a jump cut, in which Buddy moves from being left of his dad in shot 5c to being right of him in shot 6 -- we align the space according to the one grid we have, namely Rose and Mr. Hillyer. As Tversky has shown, such alignments tend to distort the space, simplifying complex spatial relations (such as "west - north west") to simple ones, (such as just "north").[100] A spatial perception of the scene then will be based on short term memory (in which only Rose and Mr. Hillyer are present), and a cognitive correction of ambiguous previous information. Thus while it is clear the zoom movement is not presented from Mr. Hillyer's perspective, based on low-level perception it seems much more like the external narrator's statement than Buddy's. But we have top-down operations at work as well; those in which we already know Buddy is the main narrator (and main subject) of the film. Therefore I would maintain that in this scene the complex and multiple modes of narrations exist side by side, on the same epistemic plane.

---

[100] B. Tversky "Distortions in Memory of Maps," Cognitive Psychology, 1981:13, pp. 407-433.

Now, the whole discussion of subjectivity/objectivity may seem like an insignificant detail to analyze, but I would like to assert that it is symptomatic of the film's portrayal of Rose. In feminist terminology she functions in the film as an object of desire who struggles to gain subjectivity (that is sovereignty and independence). And indeed, the film's plot evolves around that issue: Rose is denied the right to make decisions for herself, she is at the mercy of patriarchy (symbolized by the father and the doctor) who plan to perform a hysterectomy on her as a solution to her sexually uncontrollable nature. She is unable to save herself, but is saved by the good will of Mrs. Hillyer (the highly educated early version of a feminist, who is nearly asexual in the film, and is totally unaware of the advances Rose made towards her own husband). The film -- which may or may not have been intended as a feminist film – struggles with issues of female subjectivity and sexuality both at the level of the plot, and at the very basic visual structure.

The next scene which I would like to analyze is the one in which Rose is trying to seduce Mr. Hillyer, and fails (see appendix #2). The scene is told in its entirety from Buddy's physical and psychological point of view. Mother has left after dinner to give a lecture in town (shots #1, 3) and Dad is reading the paper in the lounge (shot# 5), while Rose is putting away dishes in the same room. Buddy and Doll, who have already figured out Rose is in love with their Dad, rush to the slightly open door of the living room to watch (shot #4). Aside from a few close-ups, most of the shots of Rose and Dad are framed through the door (#5, 7, 10, 12, 18, 20, 22, 25, 27, 31, 33). In fact, when the action moves to the couch (shot 20), and Buddy is adjusting his position as an observer, there is a (physically impossible) pan of the door, as if it was a sliding door, and the crack could be moved along an horizontal axis. The scene then, like the few others which Buddy is not participating in, is told from Buddy's voyeuristic perspective. By cutting back and forth between Buddy's gaze, and the view as seen through the partially open door, the film emphasizes Buddy as both the focalizer and the narrator.

At the beginning of the scene Doll is standing in front of Buddy, and as she is a head shorter, she is not blocking his view (see shot 8, for instance). Doll is the one who predicts that Rose is about to kiss Dad, but as soon as Rose jumps into Dad's lap, begging him to kiss her, Buddy pushes Doll out of sight (shot 11), where she remains until after the sexual action is over (shot 26). Buddy then proceeds to give Doll a verbal account of what is happening ("they are kissing"), but refuses to let her back into sight until Rose has covered herself up. Now as far as Branigan's model goes, Doll has as much epistemic information as Buddy or the audience. But surely, it is clear that there is a difference between seeing something and receiving a (veridical) verbal account of it in real time. The audience and Buddy are privileged by the film to something Doll is not – a view of the father making out with Rose. Not only is Doll denied the voyeuristic view herself, she has to rely on her brother's account. While epistemically she has all the information, she is denied the look, and is therefore judged to be in a position of (at least narratorial) disadvantage. That is, denying Doll her own low-level visual perception, and providing her with a second-hand verbal account is considered by all (Buddy, Doll and the audience) to be placing her in a position of disadvantage. But if the disadvantage is not epistemic, we need an explanation of why it is a disadvantageous position. I would claim that the reason is twofold: on the one hand we prioritize low-level visual information as being of a more veridical nature than that of verbal one. That is, we trust what we see more than what we hear, which as I have shown before has to do with the cognitive processing of verbal and visual signs. The second reason for Doll's disadvantage has to do with the structural denial of her look. Looking at other people (especially in psychoanalytic thinking) places the looker at a position of power in which the gaze objectifies the people being looked at. Looking is an active scan of the visual field in search for meaningful information. Hearing, as mentioned in the second chapter, is a much more passive sense, which is "on" all the time. When Doll is denied the view of the room she is denied the chance to act as a subject in the narrative.

The reason why Doll is not allowed to see the events has nothing to do with narrative structure, but is ideological, or patriarchal in nature. Doll should not see Rose's sexual behaviour as a role model for her own. She should not be exposed to such messages. Interestingly, the film allows Doll to *know* what is happening but not to *see* it; the film then categorizes vision as dangerous/ corrupting while high-level (rational?)

cognitive knowledge is not. Buddy, on the other hand, has to see the scene since it is a very good initiation to patriarchy. His father handles the situation well, nearly succumbs but does not, and in the end chooses his family over an affair with Rose. Not only does Buddy learn in this scene that women tend to be hysterical and irrational, he also learns that his role as a man is to have control over himself, and to stand against such behaviour. Overall, the scene asserts the role of the Father, as the one who keeps the family together, as the law-maker. And indeed, in the last shot (#37), the impressed Buddy repeats his father's line ". . .and the Persians shall not come," as if he is memorizing it for future use.

While the scene starts with both Rose and Father as objects to the kids' (and the audience's) gaze, it ends with Rose and Doll being both denied filmic subjectivity, while Buddy's (and through him the audience's) and Dad's subjectivities are reasserted. Buddy is both focalizer and narrator, so his subjectivity is granted. Dad regains his subjectivity (in the eyes of Buddy and the audience) when he gets a hold of himself, pulls away from Rose, and rationally points to her that she couldn't love him, as she said she loved Mrs. Hillyer. He then proceeds to forgive Rose, comfort her, warn her, and in short, re-assert his position of moral and physical superiority. Rose is the object of desire for Buddy's gaze, as she is never allowed her own perspective, and after the initial "attack" on Dad, she becomes very passive. Rose is also portrayed in the scene as weak (sobbing), carnal, irrational and hysterical. She is childish, and by all accounts does not behave like a mature person. All of those characteristics have been used in cinema as stereotypes which justify why women should not be seen as subjects, but as objects, or at best, helpless children.

To make things worse, as soon as Doll is allowed back in to view (shot #26), Dad starts with a speech saying: "Now a man is supposed to be a fool like this, but a woman is supposed to have some control and sense. What is the matter with you?" This verbal statement contradicts the stereotypes cinema has been promoting all along, including in this film: women are not really supposed to have common sense, only Doll needs to think that they do. Moreover, in the social environment of the 1930s, a man may have been allowed to follow his carnal desires, but a woman should have been modest and contained them, or pretend she does not have those desires. The film then sets a complex web of contradicting messages for Buddy and the audience. The visual track tells a story of overt female sexuality and lack of self-control, while the audio track asserts a norm that is clearly not met here. That is, what to the audience functions as a comical reference (we know that men really have the common sense – we have just *seen* that they do, and *hearing* that they don't will not change our mind --), functions to Doll as a educational lesson: she is not exposed to the sexual material at first sight, but the verbal message is to repress her desires. And as evidenced from her last line (shot 36), she thinks of her dad highly after the whole ordeal. Doll, as I have shown before, is therefore denied subjectivity by being placed at the bottom of the epistemic totem pole. She knows less than the participants in the scene, less than Buddy, and even less than the audience. While she knows the facts, she does not know them at first hand, and that helps in denying her subjectivity. Her perspective is both negated and subordinated to that of Buddy's (or of Patriarchy).

I am again focusing on the issue of subjectivity/objectivity because I think that this kind of cognitive reading allows for a new explanation of its functioning. Feminist criticism often focuses on the stereotypes of women, and how those stereotypes deny them subjectivity. In addition, it focuses on the feminine gaze and its threat to the main character and through him to the audience. But here we see that stereotypes and returned gaze are only part of the issue. Branigan's narratorial model shows that the narrator of any level of the text always has a subject/object relations to what is being told.[101] A subject of one level may become the object of another level of narration. For instance, when Clarice (*The Silence of the Lambs)* is recognizing the moth, she is the subject/narrator, but shortly after, the camera cuts to a shot from behind the serial killer, and in the foreground on the stove we see a gun. This shot is narrated from the implied author's perspective, and in it Clarice is an object of the narration. For Branigan, focalization is a deep (within the narrative) form of narration, since it accounts for experiencing. But as I have shown earlier, Branigan does not distinguish between visual and verbal focalization. A cognitive extension of Branigan's model will account for that. Visual focalization (particularly, but not only through point of view) grants the focalizer subjectivity. In the case of Doll and Rose this subjectivity is usually denied by higher levels of narration, but in the case of Buddy or his father it is supported by those higher levels. Female subjectivity (and thus power and control over women's own lives) is restricted here structurally, and not just ideologically. Conversely, a film like *The Silence of the Lambs* grants subjectivity to its main female character just by allowing her to focalize, and at times narrate the film. That is, the film is not necessarily a feminist film, and yet the female character has certain power and control over her life and actions, which is granted mostly structurally. And while granting subjectivity does not mean the female character cannot be at the same time objectified by a patriarchal gaze, it makes the issue of power struggle much more complex.

---

[101] Edward Branigan, op. cit. "Disparities of Knowledge" pp. 66-72.

In comparison, Buddy's subjectivity is granted from the very beginning of the film, when he is asserted as the main focalizer and narrator of the film. Buddy is nearly always present, listening and spying on other characters when he is not part of the events taking place. Only twice is Buddy not present at all, and an implied narrator takes over the narration. In both cases the parents are having arguments about Rose's future, and in both cases Dad concedes that he was wrong. The first scene takes place in the parents' bedroom late at night and ends up with a sexual innuendo. The narrative logic allows then for Buddy to be missing, although he has eavesdropped on the parents arguing about Rose earlier in the film. But the second case is much more interesting. Rose is in hospital with an ovarian cyst. In the first hospitalization, when Rose was in life danger with an extreme case of pneumonia, Buddy was present when the parents were talking to the doctor. This time he is together with his siblings outside. The doctor's muffled voice is heard in the hallway, and Buddy is even putting his ear to the door, but once the scene moves inside the doctor's office, we do not see Buddy listening again. The doctor is suggesting that while they operate on Rose's cyst, they might as well perform a hysterectomy that will cure Rose of her oversexed drive. Dad is in agreement with the doctor, but mom explodes into a speech about Rose's rights as a person, and eventually Dad is convinced that he is wrong. While the couple leaves the doctor's office as united as ever, Buddy is denied the knowledge of what was said inside. Or, if he has heard it, he has not narrated it for us. The reason here cannot be structural (as the camera could have shown Buddy listening outside), but is ideological. Buddy's initiation into patriarchy should not go as far as teaching him that it is within the power of men to take away woman's sexuality, and to treat her as a child or an animal for which adults make crucial decisions. In addition, the near hysterectomy scene is the only scene where mother puts down her foot firmly, and wins. Not only is the great patriarch wrong, but he is willing to admit it, and even to confront the doctor. Buddy's notions of patriarchy are of his father controlling his desire for Rose, not abusing his power over her. The film, rather than deal with complex and contradicting ideological positions, refrains from it altogether by removing Buddy from any position of focalization or narration, in this case addressing the audience directly. It is the only case in which the audience is in an epistemic position of advantage over Buddy, and thus he becomes the object while the audience is the ultimate subject. I would claim that the reason why the film objectifies Buddy here has to do with the desire (on the side of the filmmakers) that the film is read as a feminist text. I shall return to the issue of whether *Rambling Rose* is a feminist film, or under which conditions it could be read as such, in the fourth chapter. I hope though that so far I have shown that a cognitive reading of scenes from a film like *Rambling Rose* can shed light on feminist issues that so far have been described only in ideological or psychoanalytic terms. I believe that this cognitive reading (as an extension of Branigan's structuralist model of subjectivity) is more concrete and grounded than the psychoanalytic readings proposed so far.

Summary

In this chapter I have shown the limitations of a semiotic approach to film studies, especially when it comes to deal with visual information. This semiotic analysis treats the cinematic sign as part of a language system, and tries to articulate the cinematic experience in linguistic terms. In particular, I claimed that the photographic sign is not a good candidate for a semiotic analysis that is based on a signifier, a signified and a convention. The photographic sign is not conventionalized in abstract and arbitrary ways, and the difficulty in articulating it in such terms led theorists to complex ideological assumptions about its functioning. I then showed why a cognitive, and particularly an ecological, approach can provide valuable alternatives to the semiotic model. Using research from cognitive psychology, one can simplify cumbersome structural narratological models, and account for the specificity of the source of cinematic information (visual or verbal). A cognitive account can treat verbal information as coded, and visual information in its own specific processing manner, which is natural rather than conventional. Using the issues of narration and point of view, I followed Branigan, Currie, and others in outlining possible alternatives to existing structural models of narration. In particular, I suggested to extend Branigan's model to include sense specific information such that the model can account for the different cognitive processes at work. To illustrate my approach I used examples from *The Silence of the Lambs* and *Rambling Rose*. While this is a preliminary and limited attempt to show the value of a cognitive approach to film narration, I believe I showed the project is worthwhile undertaking in greater depth. I would like now to turn to sketching some preliminary thoughts on the value of research from cognitive science to the activity of the construction of the fabula, and specifically to the practices of gap filling and interpretation.

Chapter 4

The Act of Perceiving a Film: Interpretation and Gap Filling Practices

Introduction

In the second chapter of this dissertation I discussed the flow of audiovisual information, and how it is being cognitively perceived and processed by the film viewer. In the third chapter I discussed narrative tactics, like parallel or point of view editing, and how these tactics construct the flow of diegetic information from the film's beginning to its end. And while in both chapters I talked about the effects of perception and cognition of filmic information on the construction of the narrative, I have yet to look at the activity of the film perceiver during the viewing experience, and how that activity lends itself to the construction of the narrative. In his book *The End* Richard Neupert writes:

> The spectator requires both perceptual skills and interpretational strategies, both of which depend upon memory and hypothesis-making. In this way, the reading activities, including perception, sign identification, hypothesis-making, and filling in narrative gaps in order to guess what comes next, all become crucial for any study of how spectators interpret a film's ending.[1]

Indeed, like some other film and literary theorists, Neupert is ready to concede that the interpretation of the film is dependent upon a perceiver being actively engaged with textual information. But like many other theorists, Neupert can not (or does not) locate the exact relationship between the reading activities and the text. Instead, Neupert claims that "the spectator must proceed both by expectation and retrospection in the viewing of any film."[2] By retrospection, Neupert means the re-organization of syntagmatic textual information into paradigmatic patterns that are important to the understanding of the plot (i.e., re-organization of bottom-up information). By expectations, Neupert refers to hypotheses about what will happen in the narrative later. And while Neupert claims that both retrospection and expectation are important, he refrains from explicating the process of expectation production. I, too, will not discuss general hypotheses with regards to the narrative trajectory, but the work of this chapter will shed some light on the questions that Neupert leaves open.

---

[1] Richard Neupert, The End: Narration and Closure in the Cinema, Detroit: Wayne State University Press, 1995, P.28.

In this chapter I will look at how high-order cognitive processes -- like problem solving and memory retrieval -- enable the comprehension and construction of the narrative as a whole. As an example of these high-order cognitive activities I will be examining gap filling practices, as a moment of interaction between perceiver and text, a locus of contact between bottom-up perception and top-down operations.

Gaps operate at both local and global levels of the narrative, and they can be defined as areas where the text refrains from providing the perceiver with crucial dramatic information. Gap filling involves attention to sensory information (and lack of) as well as memory retrieval, computation, and hypotheses construction and confirmation/cancellation. It is therefore a very useful area of attention for a cognitive account of film comprehension. My analysis will show that the channel of information (i.e., image or language) has an effect on determining narratorial gaps in the first place, and consequently on the gap filling tactics employed by the viewer.

---

[2] Ibid. P.31.

But before turning to a cognitive account of gap filling, I need to review the existing theories with regards to the role of the perceiver in the interpretive process and the construction of meaning. These theories are grouped, generally, under the large umbrella of "reception studies" but as I am about to show, their focus and ideological intentions are quite varied, ranging from text oriented studies, to historical materialist approaches, and to a purely relativist postmodernist attitude. Moreover, while literary studies is heavily involved in all of these approaches, film studies has been much less enthusiastic in examining interpretation and reception. Television and Cultural studies, as well as communication models (most notably the "Birmingham school"), have mostly concentrated on the historical materialist approach, but the other areas have been significantly neglected.

Reading within the Text

One of the early influential reception studies approaches was developed in Germany under phenomenology and hermeneutics influences (see the work of Roman Ingarden, Hans-Georg Gadamer), and has culminated in the Constance School group, which included the famous theorists Hans Robert Jauss, and Wolfgang Iser.[3] While the interests of this group somewhat vary, the focus is on reception, mostly as it is inscribed by the text. I shall look here at the work of Ingarden and Iser, as they are the first to postulate moments of indeterminacy, blanks, or gaps. Ingarden, who studied phenomenology under Husserl, was interested in literature as an intentional object that becomes an aesthetic object only when completed by the reader. Unlike objects in the world, which have a particular set of properties, literary objects are described by certain schemata, but are necessarily incomplete descriptions, which retain a level of indeterminacy. In other words, different aspects of the literary object are necessarily not provided, but are imagined and actualized by the reader. Ingarden calls the "missing" aspects, textual places of indeterminacy, and the activity the reader takes in order to evade that, concretization. This concretization is very important to Ingarden, because like many of his contemporaries, his interest lies mostly in defining aesthetic value. Ingarden writes:

> The stratum of "aspects" plays a most significant role in the literary work of art, especially with regard to the constitution of aesthetic value in its concretization.[4]

For Ingarden, the literary work is an impoverished skeleton which is fleshed out (and thus receives its aesthetic value) only once concretized by a reader. But given that Ingarden's "places of indeterminacy" can apply to everything the text does not supply -- from descriptions of objects and spaces, to characters' objectives and scenarios -- the task of the reader is nearly impossible (there are infinite number of indeterminacies). Ingarden realizes that

---

[3] I am aware that Iser and Jauss emerge out of different traditions (Heidegger vs. Husserl), and am not claiming they have the same position. But for the sake of brevity and the needs of this thesis I shall talk about this branch of reception studies as a more or less unified group.

> In view of the manifold ways in which the aspects are actualized and concretized by the reader, the aesthetic apprehension of one and the same work can turn out very differently. Consequently, only some aesthetic apprehensions hit upon the work in its proper form. Even the aesthetic concretizations which are faithful to the work, can differ greatly from one another in this respect and allow very diverse aesthetically valuable qualities, and thereby also diverse aesthetic values, to appear. Here again we come up against one of the reasons for the fact that readers – sometimes even highly cultivated and sensitive critics – cannot agree in their evaluations of the same literary work of art.[5]

A few things are striking about this statement: Ingarden here tries to come to terms with the fact of varying interpretations despite his tendency towards a *proper* way in which a text should be read. That is, in his explanation the text is determinate as a system of structures (schemata), but interpretation varies because of aesthetic concretization. It is also interesting to note that for Ingarden the reading may "hit the work in its proper form" or not, that is, meaning (or lack of, through mis-reading) is

---

[4] Roman Ingarden, The Cognition of the Work of Art, trans. By Crowley & Olson, Evanston: Northwestern University Press, 1973. P. 56.
[5] Ibid. p. 62.

still guided and ultimately controlled by the text. This attitude is typical to most of the scholars of the "Constance School."

Wolfgang Iser, a scholar influenced by New Criticism whose work is within narrative theory, has developed and refined Ingarden's notion of indeterminacies. Iser was interested in further defining the relationship between the reader and the text, and he set out to show that meaning did not reside solely in the text, but was a result of an interaction between text and reader. Furthermore, the reader, who completes gaps (places of indetermincaies), becomes a participant in the process of the production of meaning. Iser writes:

> Communication in literature, then, is a process set in motion and regulated not by a given code but by a mutually restrictive and magnifying interaction between the explicit and the implicit, between revelation and concealment. What is concealed spurs the reader into action, but this action is also controlled by what is revealed; the explicit in its turn is transformed when the implicit has been brought to light.
>
> [. . .] The gaps function as a kind of pivot on which the whole text-reader relationship revolves.[6]

While Iser works on elaborating this mutual relationship, it is important to note that this relationship between reader and text re-asserts traditional hierarchies. For Iser, the text steers the reader in a certain direction, but then invites (even requires) the reader to fill in the gaps it leaves open.

---

[6] Wolfgang Iser, The Act of reading: A Theory of Aesthetic Response, Baltimore and London: Johns Hopkins University Press, 1978. p. 169.

Still, this model assumes that the reader does not and cannot impose meaning, or interpolate meaning in ways the text does not encourage or allow. The reader here is activated by the text and until then presumed passive. But historical readers come to the reading process with different (often multiple) attitudes, expectations and agendas. They may interact with the text in ways significant to the production of meaning already before encountering the first gap. Iser is interested in an ideal or theoretical reader, but in order to deal with the extratextual context he comes up with the notion of meaning and significance. Meaning originates in the text and is produced by the reader while interacting with the text, while significance is a stage "which represents the active taking over of the meaning by the reader."[7] This model opens up the room for identical meaning productions, and diverse significances, thus supposedly resolving the problem of multiple and contradictory reading practices. But Iser's tendency to construct a universal implied reader deters him from discussing ideological or other biases that affect the reader during the reading process.

Even most of the critics who are less "text oriented" and more interested in the reader's psyche and activity, refrain from relieving the text from all the burden on meaning and interpretation. Umberto Eco writes "every text, however open it is, is constituted, not as the place of all possibilities, but rather as the field of oriented possibilities."[8] And Michael Riffaterre writes that

---

[7] Dagmar Barnouw, "Review of *The Act of Reading* and *The Implied Reader* by Wolfgang Iser," Modern Languages Notes 94 (December 1979): 1209.

> The reader is the only one who makes the connection between text, interpretant, and intertext, the one in whose mind the semiotic transfer from sign to sign takes place.[9]

But at the same time

> Far from freeing the imagination, far from giving the reader greater leeway as it invites him to greater participation, reading is actually restrictive . . . [The reader is] under strict guidance and control as he fills the gaps and solves the puzzle.[10]

This discussion of the location of meaning in the literary discourse is important because it echoes the transition from a formalist/structuralist/ linguistic approach in literary studies of the early part of the century, to respectively speech act, discourse analysis, deconstruction and post-structuralism. The earlier linguistic/formalist approach treated the linguistic sign as stable, a signifier which is a presentation of a signified, an idea or meaning. The meaning of the literary (or cinematic) work of art was assumed to be located in the sign system used, and as long as one was capable of decoding the sign systems (the semiotic project), one would ultimately arrive at the meaning of the text. But Austin's model of speech act theory offered a different paradigm for communication. This model is based on an utterance, a performance of speech communication, which entails three different aspects:

---

[8] Umberto Eco, The Role of the Reader: Explorations in the Semiotics of Texts, Bloomington: Indiana University Press, 1979, p.79.
[9] Michael Riffaterre, Semiotics of Poetry, Bloomington: Indiana University Press, 1978, p.164.

[10] Ibid. P. 165.

1. The locutionary act: the act of communicating information in the normal sense of speaking. It consists of the phonetic (uttering certain noises), the phatic (uttering words), and the rhetic (doing so with "a certain more or less definite 'sense' and a more or less definite 'reference'").[11]
2. The illocutionary act: the act of force of the utterance, like asking a question, ordering or instructing. Illocutionary acts always conform to convention, and rather than a lexical convention, it is a social one.
3. The perlocutionary act: producing "certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons."[12] The consequences or effects of actually speaking cannot be pre-determined, and therefore perlocutionary acts are not conventional.

Austin writes that "(t)he truth or falsity of a statement depends not merely on the meanings of words but on what act you were performing in what circumstances."[13] The concept of utterance here is defined as not functioning in isolation, but within a performative context of a discourse. This definition is very important to reception theory. While truth-value, particularly in fictional work is a tricky matter, the idea that an utterance is incomplete without the performative (the illocutionary), draws attention to the (at least implied) reader. For Iser this is an opportunity to discuss the effect of the illocutionary on the hearer (implied reader) but also a way to substantiate the conventional practices of gap filling. But Iser always stays within the confines of a hypothetical reader, one that is constructed and activated by the text. Other theorists, however, find the implications of speech act theory to be far reaching, well beyond the text.

---

[11] J.L. Austin, How To Do Things With words, ed. J.O. Urmson and Marina Sbisa, Cambridge: MIT University Press, 1975. PP. 92-3.

[12] Ibid. P. 101.

What Austin revolutionizes most is the notion of communication, now seen as an activity based on intentions of authors, communicated through the illocutionary, and evaluated and judged by the receiver in the perlocutionary context. While the meaning of words may reside in a lexicon, the meaning of the communication act is a product of the interaction between the three levels of speech act. Moreover, the illocutionary and perlocutionary imply a certain uniqueness or non-repeatability of the utterance.[14] The modest implications of Austin's model can be seen articulated by Mary Louise Pratt. Pratt concludes that literary texts are not "autonomous, self contained, self motivating, context free objects which exists independently from the 'pragmatic' concerns of everyday discourse,"[15] but rather, like any other utterance, literary works only function in context. It is therefore not enough to account for an

---

[13] Ibid. P. 145.

[14] The unique element in Austin's model of communication was very attractive to deconstructionist and post-structuralist, and particularly to Derrida. It is easy to see that in an attempt to enlarge the gap between signifier and signified, the performative can have an honorable place in Derrida's theories. I shall not examine Derrida's argumentation with and derivation from Austin, since it is outside of the scope of this work. However, one can look at Derrida's "But, Beyond . . ." Critical inquiry, 13, (1986), 155-170. See also Robert Scholes _"Deconstruction and Communication "Critical inquiry, 14 (1988), 278-295.

implied reader, instructed by the text, but one need to turn attention to the context of reading, and to the functions of literature in various social contexts. But the locutionary and the perlocutionary still postulate an agent who has intentions to achieve results, (preferably an action) through the utterance and the communication act. And while it is possible to imagine the literary character, narrator, or even the implied author performing such agency, or having such intentions, it is hard to ascribe such strong intentionality to film. Even if one wants to discuss the implied filmmaker, and his/her intentions (as Currie does, see chapter #3), one has to tackle the fact that film conveys and performs at best a very diffused form of agency; one which is comprised of lights, camera angles, framing, camera movements, sound recording facilities, action, choreography, music, and many more aspects which are used by the filmic storytellers. And as I have already shown, these different tracks of information may tell different or even contradicting stories, a tactic which is regularly used in film narration in order to encourage different hypotheses about what is going on. In other words, while Austin's model is useful, and the idea of communication acts opens up the door for an active "reading" on the side of the perceiver, it needs to be carefully applied to film, and the question of intentionality (especially the one based on agency) needs to be addressed before any application will be coherent and valuable.

---

[15] Mary Louise Pratt, Toward a Speech Act Theory of Literary Discourse, Bloomington: Indiana University Press, 1977, p. 115.

Reading in the Reader

Some reception studies approaches have adopted a much more radical interpretation to the context of reading, and the derivation of meaning while reading a literary text. Originating from a Marxist materialist framework, these theories focus on the place of ideology and politics in the reading practice. This approach does not try to construct an ideal or implied reader, but to research actual readers, and to give an account of the variety of readings historically evident for singular texts. Initial work (particularly the Birmingham school) was done with easily identifiable groups such as gender, ethnicity, class, race, and sexual orientation. The group has been assumed to be unified and idealized to represent alternative or oppositional perspective, but a complex intersection of variable elements of identity was not proposed. Indeed, if one goes all the way with this historical approach, one may find it very hard to come up with conclusions that go beyond individual interpretation. While this may be a valid empirical goal, it would be nevertheless, quite useless theoretically. But reception theory is not only historical and empirical. It also postulates a very different launching point for literary studies. Mark Silberman writes:

> Reception theory specifically displaces the focus from the literary work of art to the reader, to the receiver who constitutes the text in a historically mediated process of reading. [. . .] By defining the reader as the source of meaning, reception theory reformulates the problematic relationship between past and present, and between literary history and extra-literary or pragmatic history. In short, it

> opens the door to rewriting literary history and to redefining the literary canon.[16]

Indeed, by focusing on the ideological, political, economic, and psychological elements of the reading process, reception theory not only changes the focus for literary studies, but also locates meaning or at least interpretation[17] not in the text, but in the activities of reading. For some, (most notably Stanley Fish and Norman Holland) interpretation is perception. Fish writes:

> The sentence is no longer an object, a thing in itself, but an *event*, something that *happens* to, and with the participation of, the reader. And it is this event, this happening – all of it and not anything that could be said about it or any information one might take away from it – that is, I would argue, the *meaning* of the sentence.[18]

But others try to separate perception (the event of reading) from interpretation and meaning. George Dillon, for instance identifies at least three levels of reading: perception, comprehension, and interpretation.[19] Perception refers to the recognition of words and sentences. In comprehension one places this linguistic chain into a frame of reference, a literary code, or a schemata of expectation. Interpretation is a stage in which the reader relates her understanding of what is going on to the (assumed) narratorial intentions. In this way "interpretation governs comprehension and perception in that we tend to see what we have inferred the writer wants us to see."[20] Incorporating the structuralist models into a theory that attributes the power to the reader, but still uses the traditional formal/linguistic views to describe perception, this approach merely shifts the weight towards the reader, but does not propose a new approach to reading altogether. This position is typical to many reception studies theorists. Janet Staiger, for instance, claims that

---

[16] Mark Silberman, "Review of Holub's Reception Theory," New German Critique 33 (Fall 1984), P. 250.

[17] The debate over reading, interpretation, comprehension and meaning, is, as expected quite convoluted. I shall not elaborate on that debate since it is a subject for a whole other thesis, but will use a few of the positions, as a lead way towards my cognitive arguments.

[18] Stanely Fish, "Literature in the Reader: Affective Stylistics," in (ed.) Jane Tompkins, Reader response Criticism: From Formalism to Post Structuralism, Baltimore: Johns Hopkins University Press, 1980, p. 72. Emphasis in the original.

[19] George L. Dillon, Language Processing and the Reading of Literature: Toward a Model of Comprehension Bloomington: Indiana University Press, 1978.

> Thus, a reader interpreting any work of literature or film will be drawing upon interpretive frames historically available to him or her, and these frames will be influential even in the act of perception or the process of comprehension and evaluation.[21]

Staiger admits that in the reading process "controlling conventions, linked to ideologies, win out over illusionary variety,"[22] but the potential for indefinite number of readings is theoretically possible. The reason, according to Staiger for why most readings are quite uniform is that readers have been socialized and ideologically manipulated both in literary traditions (genres, and dramatic conventions) and extra literary environment (capitalism, patriarchy). And while there is no doubt that ideological biases would affect and alter the reading process, often providing different interpretation and personal significance, I would be reluctant to free that process from the text. As I have shown in chapter 3, some narratorial constructs, such as point of view for instance, imitate and thus trigger biological mechanisms that operate prior to high-order cognitive (both structural and ideological) ones. I would like to suggest that some perceptual and cognitive mechanisms guide our interpretive processes and lead them to be generally uniform (for instance, identifying with a main character because of point of view structure that is established). This uniform interpretation is thus not ideologically based, but textually guided and encouraged. Now the cognitive aspects of a reading may lead to an ambivalent reading, one that raises ideological issues, and therefore requires the perceiver to use her own ideological judgement, but the ideological stage is here secondary to the cognitive one. For instance, it is possible to argue whether *Rambling Rose* is a feminist text. A cognitive analysis of the scene in which Rose jumps into Mr. Hillyer's lap shows the film actually provides the perceiver with conflicting messages. On the one hand, the image track tells us that Rose is infantile, hysterical, and generally reduced to an incoherent subject. But the verbal narrative tells us that women generally have more sense than men do. The ambivalence of the text leads to a complex reading. If one is pre-exposed to a feminist agenda, one may critique the film for reducing Rose to an uncontrolled childish id (prior to the development of the ego). But if one is not aware of the ways in which Hollywood cinema has

---

[20] Ibid. P.xx.

[21] Janet Staiger, Interpreting Films: Studies in the Historical Reception of Cinema, Princeton: Princeton University Press, 1992, P.21.

---

[22] Ibid. P.10.

tended to infantilize women characters, one may be impressed by the amount of slack that Mr. Hillyer (and the film) is cutting Rose, and consider the film to be asserting female sexuality in positive terms. The point here is that the ideological reading is present and essential for interpretation, but is engaged by a primary bottom-up cognitive reading; that is by perception of images and sounds that amount to actions, that only then lead to comprehension and interpretations. Moreover, while the interpretation of the text as either feminist or not is open for debate, the plot elements (who tried to seduce whom) are not. That is, it is clear to all viewers that it is Rose who is trying to seduce Mr. Hillyer, and the ambiguity resides in what is the implication of the scene-- not so much to the film, but to the issues of the representation of women's sovereignty over their bodies and selves. To put it in Iser's terms, the meaning of the text is unambiguous (action-wise), but the significance depends greatly on the actual historical positioning of the individuals in the audience. It is possible that some ideological positions may alter our sensitivity to textual information, and thus affect our perception. But I would intuit that it wouldn't affect primary perception (i.e., the biologically-oriented perception of basic power relations, basic action trajectories, etc.). Either way, I believe that the ideological aspects of interpretation work in tandem with the cognitive aspects, and together form the reading of the text. Put another way, we can use David Bordwell's terminology of comprehension and interpretation. For Bordwell,

> Comprehension is concerned with apparent, manifest, or direct meanings, while interpretation is concerned with revealing hidden, nonobvious meanings.[23]

Comprehension covers referential and explicit meanings, references the viewer extracts from the text that could be considered literal meanings (the diegetic world, or direct metaphors such as the scales of justice). Interpretation on the other hand, covers implicit and symptomatic meanings, meanings that the viewer derives out of positing problems, questions, examining themes, or speculating about the intentionality of the author.[24] In other words, comprehension refers generally to bottom-up processes of perception, identification, and to some degree the high-order cognitive operation of causal re-organization of the fabula. Interpretation, on the other hand refers to high-order cognitive operations only; those of postulating questions of intentionality, ideology, symbolic implications, irony, and other meta-textual and contextual issues. An analysis of the above mentioned scene from *Rambling Rose* according to this model would determine that the referential and explicit meanings (i.e., comprehension) are easily discernible, but the fact that visual and aural information are at odds can only be explained in an implicit or interpretational model. Bordwell goes on to say:

---

[23] David Bordwell, Making Meaning: Inference and Rhetoric in the Interpretation of Cinema, Cambridge, MA: Harvard University Press, 1988, p.2.

[24] Ibid. pp. 8-9.

> Still, in this book I will not be much concerned with comprehension. My stress here falls on interpretation, conceived as a cognitive activity taking place within particular institutions.[25]

Much like the materialist reception theorists, Bordwell here is nearly ready to discard the text in order to concentrate on high-order cognitive activity, which according to him operates in particular institutions that are independent of the text. But once he declares the separation between comprehension and interpretation, and decides to focus on interpretation only, Bordwell loses ground. As George Wilson shows, Bordwell's model assumes the implicit and explicit aspects to be necessary conditions for interpretation (which in his model is based on intentionality), but he refuses to address that connection.[26] But as the scene above so clearly shows, the implicit reading, and search for the feminist (or anti-feminist) cues in order to have an interpretation of the scene can only be meaningful if the modes that lead to comprehension are taken into account. That is, the tension between the visual and audio track and how they are perceived respectively and concurrently lead not only to Bordwellian comprehension, but affect interpretation as well. And as much as Bordwell's project is illuminating and important, he still reiterates the old division between structural (textual) approach and the reception (reader-based) one.[27]

---

[25] Ibid. p.10.

[26] George Wilson "On Film Narrative and Narrative Meaning," in Film Theory and Philosophy (eds. Richard Allen and Murray Smith), New York, NY: Oxford University Press, 1997, pp.224-225.

To sum up, textual-oriented reception studies (such as Iser's) differ from the historical approaches (such as Staiger's or Fish's) in the following ways: while textual approaches look at an object (text), the historical approach looks at an event (perception); while the former looks for meaning, the latter examines functions; and finally, while the textual group constructs a universal reader, the other group looks at an actual historical one. We can sum it up by saying that for Iser, the text and its perception determine the interpretive process, while for Staiger, ideology determines perception and interpretation.

In cognitive terminology, we can posit this debate as one between top-down emphasis (beliefs determine perception) to bottom-up approach (textual information leads to higher cognitive activity such as interpretation). And given that cognitive science research has a lot to say about top-down/bottom-up operations, I believe it can shed some light on this debate. I shall look at gap filling in cinema as a particular point in which textual information is missing, thus inviting the viewer to fill in gaps, and will examine how these gaps are determined and filled. I believe that this close cognitive look will reveal that gap filling, and interpretation (which is a necessary condition for gap filling), are complex operations involving both low level perceptual material and high-order cognitive activities. But before moving on to discuss cognitive gap filling,

---

[27] I find this professed preference of the book on "interpretation" (in that particular way that Bordwell defines it) to be quite ironic, as Bordwell's orientation is more of a neo-structuralist than a reception theory direction, and his cognitive approach is harnessed towards that end. Even within Making Meaning he cannot sustain a consistent argument favoring top-down cognitive mechanisms over bottom-up textual cues.

I need to discuss in greater detail the nature of gaps, and provide an initial taxonomy of gaps and how they are constituted in our sense experience.

What are Gaps?

Whether they employ a top-down or a bottom-up approach, most literary and film critics today agree with the deconstructionist and post-structuralist claim that the text is at times incoherent or even contradictory. That is, even if an ideal and competent reader performs the reading as instructed by the text (under the most conservative account of reading practices), the text itself at times offers multiple organizing patterns that may compete or be logically inconsistent with one another. Coherency, it is suggested, is something that the reader tries to achieve out of fear of logical contradictions. James Kincaid says:

> The reading of literature is in large part a search for the organizing patterns [. . .] that will make coherent all the numerous details or signals we pick up along the way. Readers proceed with the assumption that there must be a single dominant structuring principle.[28]

The reason why the reader resists the notion that the text is incoherent, and tries at all cost to resolve the internal contradictions and to find the organizing patterns, is, according to Kincaid, biological, based on a reflexive desire to resolve all crises. Kincaid cites Ralph Rader who hypothesized that readers have an evolutionary need to make decisions when confronted by ambiguities.[29] In addition to the tendency towards coherence, Ulrich Neisser notes that adults in the Western world construct events of an experience into a temporal sequence.[30] Psychological research on eyewitness narratives of everyday experiences support this claim.[31] The tendency to linearize into a cause and effect sequence, can explain why (at least) Western film viewers tend to prioritize taking syuzhet material (plot) and transforming it into a linear story, a narrative (fabula). The tendencies for coherence and linear temporality are challenged when the film provides a gap. A gap represents lack of spatial, temporal or dramatic information and as such invites the perceiver to fill it for the sake of coherence.

---

[28] James R. Kincaid, "Coherent Readers, Incoherent Texts," Critical Inquiry 3 (Summer 1977): 783.

Gaps may vary in nature as some are important to the narrative and some are not; many gaps are inevitable and negligible. For instance, a film rarely shows an entire space within a scene; even when the events are covered from multiple perspectives, there are sections of the space that are not shown.[32] The viewer, though, "completes" the space in her mind, to the point that she is not aware that she has not been shown the entire space. This spatial expansion is done primarily by bringing in knowledge (schemata) about the physical world, and about continuation of spaces beyond our visual field. We also use our knowledge of conventional spaces to fill in visual spaces that are missing or very briefly introduced. In cognitive psychology research, subjects were shown a picture of a space for a brief time, and then asked questions about it. In the case of an office space, many claimed that it was furnished with books on bookshelves, even though it was not. The subjects here were using an "office" schema (much like Minsky's frames and Schank and Abelson's scripts, see chapter 2), and completed (or replaced) information about the space -- information that wasn't actually provided -- with prototypical patterns. Similarly, in realist drama we complete spaces beyond the borders of the frame, and we even complete spaces that have been presented on the screen for too brief a time to even perform a complete visual scan, let alone a meaningful comprehension. In contrast, sci-fi and horror films often refrain from providing an establishing shot, instead concentrating on close-ups in dark, hazy, and visually restrictive situations, since much of the suspense is built on disorienting the viewer in unfamiliar, un-completable spaces (see the discussion of *Alien* in chapter 2). Such films actually deprive the audience of the possibility of speculating about the spaces they are watching, by making the spaces unlike any experienced by humans (or previously experienced on film).

---

[29] Ibid. P. 785.

[30] Ulrich Neisser, Cognitive Psychology, New York: Appleton- Century- Crofts, 1967, P. 290.

[31] See Siegfried Spence, Ludwig Malpass and Roy S. Kuehnken (eds.) Psychological issues in Eyewitness Identification. Mahwah, NJ: Lawrence Erlbaum Assoc., 1996.

[32] In fact, in interior scenes, in order to maintain continuity of lighting one wall (or one side of the room) is usually used just for lights, and unless a window or another light source is explicitly shown, that direction may never be seen at all.

Sound effects, too, help expand space beyond the limitations of the visual frame. Some sound effects emanate from the visual field, but many are outside it, giving the audience an indexical reference as to the spaces and events outside the visual field. The whistle of the kettle, a knock on an invisible door, and voice over dialogue (that is not an interior monologue), all indicate to the viewer the layout of a space. Sound effects may also define whether an interior space is located in rural or urban setting, and other important dramatic information. In *The Silence of the Lambs* for instance, the cries of the dog indicate to the killer and the audience at once that the girl has managed to seize his dog (appendix 5, shot 7a), and the doorbell ring shortly afterwards indicates that the FBI has arrived (shot 31). It is interesting to note here that sound effects provide an intriguing cognitive phenomenon. While sound effects are heard and therefore perceived aurally, they designate spatial properties as much as events, and we may note that in some cases they are stored in memory as images, and not as coded propositions. Put differently, since sound effects epistemically expand the visual field beyond the frame lines, I intuit that SFX sometimes stand in for spatial visual information. The cross path from audio perception to high level visual cognition is fascinating, and deserves attention, but is beyond the scope of this project. Whether spatial or aural (sound effects), these gaps are ever present and unavoidable and are filled in by inferences the viewer makes, based on her knowledge of the natural world.

Temporal gaps are of a different category altogether. There is always a relationship between the duration and pace of the story (fabula), and the one of the syuzhet. Gerard Genette called the discourse (syuzhet) treatment of temporality "duration" (later he changed the term to speeds). Duration exists in relations of descriptive pause, scene, summary and ellipsis, to fabula events.[33] The "descriptive pause" refers to moments of plot events that are stretched in the film discourse. For instance, in *High Noon* (Zinneman, 1952), the plot starts at 10:30 and ends at noon. The film's duration is also 90 minutes. However, the first hour of fabula time is condensed to about 30 minutes of syuzhet, while the last few minutes of plot (the showdown) are stretched in a descriptive pause that lasts nearly 20 minutes of discourse screen time. "Summary" refers to scenes like the breakfast estrangement in *Citizen Kane*, and a scene refers to a relationship in which story time and discourse time are equal. Ellipsis, on the other hand is a case in which fabula time has passed, but the syuzhet refrains from representing it. In other words ellipsis are narratorial gaps. Some temporal ellipses behave much like the spatial gaps; they are easily filled, and consist of mostly mundane information that has no dramatic importance (sleeping, eating, showering, etc.).

But some temporal gaps are very important to the narrative, as the order of syuzhet vs. fabula information determines what the perceiver can and does know about the fabula. To begin with, not all films are told linearly (i.e., starting at the beginning of the fabula, moving forward towards its closure). Some films, as in the detective genre, start after a crime has already been committed, and progress to examine the past (through flash backs, interviews, and other devices), as well as move into the forward, linear, toward-the-future progression of the film's present tense. Genette calls the narratives in which the syuzhet is out of sequence with the fabula "anachronies." Anachronies can refer to the past (analepses) or to the future (prolepses), although this flashforward type is rare. Ellipses, according to Genette, complicate two aspects of the temporal structure (in Genette's terminology, tense) of the narrative, that is order and duration. The perceiver then first needs to re-arrange the order of the scenes, and then "fill in the blanks" or hypothesize as to what happened. The detective genre is a perfect example since the perceiver (usually together with the detective) is engaged in piecing together the causal order of the events, but also in predicting where they can get answers to their questions, or what questions to even ask. In other words, the detective genre is premised on the detective and perceiver coming up with hypotheses as to who committed the crime and why, so that they can have a "direction" of investigation. In *The Silence of the Lambs,* Clarice is looking at a dress that one of the victims was sewing for herself, and she realizes that all the victims were large size young women. The camera zooms in on the dress and when it cuts back to Clarice, we know that she has come up with a new hypothesis about the killer. This hypothesis (that the killer is making a suit made of human flesh) is conveyed to her superior verbally in the next few shots.

---

[33] Gerard Genette, Narrative Discourse: An Essay in method, Ithaca: Cornell University Press, 1980. See also, Narrative Discourse revisited, trans. By Jane E. Lewin, Ithaca: Cornell University Press, 1988, pp. 33-37.

Temporal gaps in narrative films may be temporary, and once the film viewing is complete and the narrative has been constructed, all pieces of the puzzle are provided. Such is the case in a film like *Vertigo* (Hitchcock, 1958), where at one point the film presents the audience with a gap, but eventually, the film itself fills it. After Madeleine's death, the audience and Scottie (because of his fear of heights) are left to wonder whether Madeleine jumped from the bell tower top, or whether she accidentally fell in the intensity of the moment. But later the film reveals to us that Madeleine was actually pushed to her death, and that the woman who both Scottie and the audience thought of as Madeleine, was actually hired by the husband/killer to act as if she is Madeleine as part of the assassination plot.

But dramatic gaps may not be fulfilled explicitly by the film, although the hypothesis developed by the audience is subtly encouraged. In *Dead Poets Society* (Peter Weir, 1989) for instance, the main character commits suicide. The film never shows the suicide, and while the audience knows something has happened, it takes long minutes before the film affirms the death hypothesis.[34] The suicide, the climax of most of the dramatic layers of the film, is never shown, and is usually only referred to in other contexts. Yet the hypothesis that it was a suicide and not an accident is highly encouraged by the film, and I shall return to analyzing the gap filling practices of this scene soon. Dramatic gaps may also affect the entire closure of a film, causing either story, text, or discourse to remain open in different ways. While this phenomenon is more likely to occur in European "art" cinema than in mainstream Hollywood, some films like *Chinatown* (Roman Polanski, 1974) leave at least some tracks of the plot open by not filling some gaps established early on in the film.[35]

---

[34] A detailed analysis of the scene is provided bellow.

[35] For a comprehensive review of issues surrounding closure, see Richard Neupert's book The End. Ibid.

Some gaps may be read as such, but are actually not informational gaps according to the definition provided above. In these cases the perceiver is under the impression that she is coming up with hypotheses about missing information of her own accord, but in fact the film provides the information for that hypotheses, but it does so in a channel she is unaware of. I shall call those cases illusory gaps. An example which I shall analyze in detail below is a scene from *Dangerous Liasions* (Stephen Freares, 1988) in which subtle visual information affirms what seems to be a narratorial gap. Given the genre of the film (drama), the nature of gaps has to do less with action and more with emotions. Moreover, the film positions the audience in a privileged place as they know what both main characters are plotting together and apart. Indeed, much of the enjoyment of the film is based on the fact that the perceiver knows more than most characters at any given time. In the discussed scene though, the question is how an "un-collaborative" character, Madame de-Tourvel, feels towards one of the main characters, Valmont. A hypothesis is set in motion, but as I shall show, the film provides that information visually, so a real gap never occurs. While the audience may not be aware of that information, it is as effective in influencing the comprehension of the narrative. I will claim that illusory gaps are a result of multiple channels of perception (i.e., visual and audio), which leads to complex cognitive processing, one that affects the overall understanding of the film. Robert Burgoyne writes that:

Illusory gaps seem to support Thomas Elsaesser's claim that:

> ellipsis exhibits a strong mark of enunciation, but, unlike other enunciative systems such as montage, it usually hides the mark of enunciation, and lets the reader feel he or she is making the connections.[36]

And although I believe that generic plausibility, general knowledge of the world, as well as other factors do affect our gap filling operations, I still agree with Elsaesser that those are strongly guided by the cognitive information that is readily available for us in the text. I shall now turn to discuss how we cognitively fill in gaps.

Cognitive Gap Filling in Theory

The perceiver of a text, upon encountering a narratorial gap is required to devise a strategy for how to fill the gap, and make the text cohere. This strategy generally involves coming up with hypotheses as to what is likely to have happened, hypotheses the perceiver is hopeful are correct. In logical terms the process of coming up with inferences to the best explanations is called abductive reasoning. Abductive reasoning not only deduces a set of possible worlds that could exist based on the facts we have (i.e., in our case different narratorial hypotheses) and on axioms, but also provides a way to codify the preference of one model over the other.

But until the very end of the text's transmission (or consumption) the perceiver is aware that some hypotheses may change, and that the narrative may still be different than what the perceiver assumes during the perception. Importantly, then, the perceiver is ready to alter, cancel, or embrace new hypotheses as the text provides her with new information. This process is very similar to formulations of belief revision (a sub-field of nonmonotonic reasoning), in which a logical theorem is challenged by new information and forces "backtracking," the introduction of new inferences, new conclusions, and revision of previously drawn theorems.[37] In other words, the fabula, or the complete story is a product that the perceiver commits to only after the perception of the text is over. And while the narrative as a product is being constructed during the perception, it is constantly in flux, or open to be in flux, until perception is over. As such, the conclusive narrative of a text is a post-perception product. And as a post perception product, the narrative is constructed from memories, which are re-organized in a causal order so as to yield the most coherent story possible. It is important to notice then that the story or fabula is a product of an array of high-order cognitive activities, ones that are significantly different than the low-order perceptual processes. Research on the empirical evidence of narrative structure supports this conclusion. In a series of related experiments conducted by Gee and Grosjean, subjects were asked to read and then recount a short narrative. The spontaneous pause duration between sentences were analyzed and then matched with Lehnert's complex analysis of narrative structure into simple plot units.[38] What Gee and Grosjean found out was that

---

[36] In Robert Stam, Robert Burgoyne and Sandy Flitterman Lewis, New Vocabularies in Film Semiotics, New York & London: Routledge, 1992, p. 120.

---

[37] C. E. Alchouron, P. Gardenfors & D. Makinson, "On The Logic of Theory Cahnge: Partial Meet Functions for Contraction and Revision," Journal of Symbolic Logic 50, 1985, pp. 510-530.

> [. . .] as the narrative complexity of a break between two sentences increases, the pause produced by a speaker also increases – and in a very systematic way.[39]

But while Gee and Grosjean were primarily interested in providing empirical evidence to narrative structure, their research reveals another interesting phenomenon. The correlation between story parsing and pauses was found only when subjects re-told the story after reading it to themselves. When they were asked to read the story aloud (even at a second reading), the pauses did not match the narrative structure parsing so well.[40] Gee and Grosjean do not explain why spontaneous re-telling reflects so much more accurately the narrative structure, but in the context of our discussion it is clear; re-telling takes into account that the narrative has been fully comprehended and interpreted before it is re-told. It is a post perception activity, and as such, the performer (the former perceiver) has a full concept of the narrative. Reading out loud, on the other hand, does not provide the correlation between pauses and narrative complexity, because it is hard for the perceiving agent to anticipate accurately where narrative units begin and end (or where pauses should be placed). According to this explanation, the correlation between re-telling and narrative complexity is easily understood, as both require a "post" perception moment in order to be fully actualized.

---

[38] James Paul Gee and Francois Grosjean, "Empirical Evidence for Narrative Structure," in Cognitive Science 8, 1984, pp. 59-85.

[39] Ibid, p.72.

Gaps configure in interesting ways into this explanation of narrative: they provide a moment where the perceiver is invited to produce a hypotheses, but is also reminded that she may be wrong, or that the text may be misleading her. It is only after the transmission of the text is over, that the perceiver can re-organize all the data, determinately fill in some gaps, and confidently claim that others shall remain open. But saying that the narrative (as a product) is a post perception result, does not mean that the construction of the narrative (as a process) happens entirely in post perception or in high-order cognitive arenas. The narrative is constructed as perception occurs, and it is constructed based on bottom-up perceptual activities, as much as by top-down impositions of belief, ideology, and expectations. Cognitive film theorists have struggled with the issue of perception and interpretation, and in the next few paragraphs I shall provide a short summary of their positions.

Edward Branigan provides an adaptation of Alan Williams' formulation that

> [. . .] when we watch a narrative film, we are actually watching four different films: a celluloid strip of material; a projected image with recorded sound; a coherent event in three dimensional space; and finally a story we remember (i.e., the film we think we have seen). There are perceptual "gaps" between each of these four films in

> which certain facts are concealed and "forgotten" about one film in order to perceive another.[41]

Branigan here describes the phenomenon of ignoring editing within a scene -- which implies spatial and often temporal skips – in favor of accepting the dramatic unity the scene conveys. Out of the same principles of dramatic coherence we ignore jump cuts and other visual and audio inconsistencies, which, according to this account we perceive, but dismiss for the overriding needs of the fabula. Branigan here accepts a modular approach to the construction of the narrative, one which fits nicely with his notion of narration since it evolves around hierarchical structural levels of both cognition, mind, and narrative.[42] Branigan claims that "comprehension proceeds by canceling and discarding data actually present, by revising and remaking what is given."[43] What is striking about this account is that it assumes that this process (comprehension) works in one direction, from top-down impositions (since the needs of the narrative are computed in high level cognitive procedures) onto the bottom-up perceptions. And while this account of comprehension is probably true much of the time, it should not be presented as the only possible scenario. In some cases, for instance, the cinema of Godard, jump cuts and other disruptions to cinematic conventions force (via bottom-up perception) a comprehension (or in Bordwellian terms an implicit interpretation) that the film is an essay about conventional cinema. The opening shot of *Breathless* (Godard, 1959) shows a close-up on a page from a newspaper. A close voice over of a man mutters some words, and the conventional assumption is that he is responding to what he reads on the page (i.e., that it is a point of view shot). But the camera then starts tilting up on the page, and continues up to show the face of Michel, revealing that we were seeing the back of the paper, thus canceling our hypothesis that it was a point of view shot. In other words, low level perception of visual information first encouraged a high level hypothesis (p.o.v shot) and then cancelled it when Michel's face was revealed. Now this moment is quick and probably does not produce a complete interpretation of theme, but it disorients the audience enough, and indicates that this film is not about to be a conventional film, which is a high-order implicit realization. In contrast, the last scene of a conventional mainstream film *The Usual Suspects* (which I described in chapter 3), changes the terms of narration, and the relationship between film and viewer in its last few minutes. As many crime films do, throughout the film, it encourages hypotheses about the events in conventional ways. But these conventions are broken and mocked in the last scene, disabling all the hypotheses produced so far, without offering another explanation as to what actually happened. In the last scene what is given, the information presented, (in close-ups of objects, photos, text, etc.) is the source of the change in the state of all previous hypotheses, as everything that preceded, (not only hypotheses, but events visually present as well) is marked as a fabrication. Branigan's model of hierarchical comprehension does not leave room for such filmic interruptions in the perceiver's top-down operations, or at least these interruptions are always enveloped within high-order narratorial and perceptual structures. And if for Branigan the comprehension of existing data could be readily written off, or revised, so much more is the danger with gaps, or moments of indeterminacy, which do not provide information at all. Although Branigan does not provide a comprehensive description of gap filling he does refer to it occasionally:

---

[41] Edward Branigan, Narrative Comprehension and Film, New York and London: Routledge, 1992, p.84.

[42] Branigan here cites the work of Ray Jackendoff, Andy Clark, Jerry Fodor, Howard Gardner and Marvin Minsky. As I have shown in chapter two, these researchers agree very little on the architectonics of the mind. But Branigan here does not adopt a particular model; rather he is concerned with implying the importance of cognitive science to the understanding of the interpretive process of film, without getting into the different positions of these debates, or without even making very concrete claims about how it operates in film comprehension.

[43] Ibid, P.83.

> By conceiving of narration as a type of verbal (and imagistic?) description offered by a spectator, one is, in effect, analyzing *interpretive* statements. One is mapping a course of thought, the use of language, rather than discovering the absolute properties of an object or discovering "cues" that are "in" an object – the text objectified. Interpretation thus construed exhibits something of the nature of an explanatory "theory." Interpretation in this sense includes the "filling in" of certain data (from the top-down) which seems to be "missing" at some moment in the text as well as the construction of macro-propositions which are *about* the text though not strictly *in* it, or denoting it. Structures that are achieved in cognition cannot be reduced to a list of phenomenal forms or cues. We demonstrate our knowledge of narration, of "how to go on," by interpreting, by going on.[44]

---

[44] Ibid, p.112.

Gaps then, given that they are never "in" the text, are assumed to function only in high-order cognitive constructions of the text, one that are filled "from the top-down."

What I caution here against is the tendency, which both Bordwell and Branigan share, of separating top-down operations from bottom-up perceptions, and prioritizing the former without allowing for theoretical moments in which this dominance is broken. As I have shown above with both *The Usual Suspects* and the *Breathless* examples, the assumptions about this top-down dominance lead to a partial and reductive description of what actually happens during viewing. But even more importantly than the examples which contradict this model of dominance, I would like to point out that some of the most intense moments in our cinematic experience are moments when a film surprises us from the bottom-up. In these moments bottom-up perception interferes with "top-down" assumptions, forcing the viewer to re-assess previous information, and to come up with new and revised logical models of causality and implications.

Gaps can be generally divided into two groups, implied and necessary. Implied gaps (also known as blanks) are gaps that are not essential to achieving dramatic coherence (like assuming that the characters sleep, eat, and use the bathroom throughout the temporal duration of story time). Dramatically necessary gaps are gaps that necessitate inference production in order for the plot/text to cohere and make sense. Chaffin suggests that when we encounter a necessary gap we engage in producing bridging inferences, or what I earlier called gap filling.[45] In addition to implied and necessary inferences, Chaffin identifies a third group, that of invited inferences. When encountering the sentences "the bird is in the cage" and "the cage is under the table," the reader can come up with the correct inference that the bird is under the table. This inference may not be necessary for the understanding of the text, but can occur, or is invited by the text.[46] The interesting phenomenon about invited inferences is that they are likely to be remembered as provided by the text, that is remembered as actual bottom-up information, even though they were inferred rather than experienced. Keenan and Kintsch, for instance, showed in recall tests that the longer the interval between exposure to the text and answering the questionnaire (up to 20 minutes), the more likely it is that the subjects thought they heard the last sentence ("the bird is under the table"). Keenan and Kintsch explained this phenomenon by speculating that surface structures like syntax are lost or discarded from memory faster than content is.[47] Invited inferences then are inferences that are causal and logical and consistent with information provided, but are not perceived as such. They are remembered however as perceived information because the content was (correctly) not categorized as an hypothesis. The phenomenon of long term memory concentrating on content rather than on surface structure may explain why out-of-frame sound effects in film are sometimes remembered as visual spaces, which is a cross-modal cognitive operation. At least within a scene, the SFX that expand the visual field (like a doorbell, car honk, or the kettle) – are remembered as spatial information, that is, as a floor plan indicating that the kitchen is to the right of the living room, while the door to the house is to the left. Sound, action and editing here are computed to create a grammar of spatial continuity within one scene. Much like Piaget's subjects (see chapter 2), we learn the filmic space of a scene by stable landmarks (couch, door, etc.), and the relationship of camera/character to these landmarks. But from this fixed stage of coordination by landmarks, the perceiver may attempt to construct an abstract cardinal system, one that provides with a floor-plan of the space. And given that this floor-plan is abstract and propositional, it is likely to be stored in memory as such, regardless of the initial source of perceptual information. It is important to note that this model of invited inferences of spatial arrangement works within a scene (which is defined as a dramatic unit occurring at the same time and same space) and not across scenes, which operate beyond temporal and spatial boundaries.

---

[45] Chaffin R. "Knowledge of Language and Language About the World: A Reaction Time Study of Necessary and Invited inferences." In Cognitive Science 1979, 3, pp. 311-379.

[46] Ibid.

Invited inferences are also those gaps that could be filled by a simple application of a frame, a schemata, or a script theory.[48] In those cases generic expectations -- both from artistic genres such as Westerns, or thrillers, as well as from everyday experience – help us fill in gaps. In a

---

[47] Keenan J.M. & Kintsch W. "The Identification of Implicitly and Explicitly Presented Information" in W. Kintsch (ed.) The Representation of Meaning in Memory, Hillside, NJ: Erlbaum, 1974.

[48] See the discussion of the work of Marvin Minsky and Schank and Abelson in the second chapter. See also M.L. Grick & K.J. Holyoack, "Schema Induction and Analogical Transfer" in Cognitive Psychology, 1983.

Western then, we expect to see a showdown, and when that camera "enters" the saloon, the audience can "complete" the unavailable parts of the space with a known gallery of characters and objects. Similarly, once we recognize a scene takes place in a kitchen, for example, we expect certain objects to be present. In other words, we employ a script or a header of a "kitchen" or a "saloon" scenario, and we then fill-in the invited inferences from the top-down.

The difference between invited inferences and necessary ones is striking though. While invited inferences are processed as actually perceived information, necessary ones slow down comprehension, while subjects search for plausible bridging mechanisms. In research on gap filling and comprehension Havilland and Clark found that sentences like "the murderer was one of John's friends" was read slower after "John died yesterday," and faster after "John was murdered yesterday."[49] The gap between death and murder requires re-organization of previous data, and therefore slows down the comprehension of the text. More recent research shows that the slowed down response has to do with memory activation. While we are engaged in reading, comprehension practices always attempt to maintain coherence on two levels: local, and global. Local information is matched with immediately preceding information, which is mostly available to short term memory (7± 2 items).[50] The maintenance of global coherence, on the other hand, requires the reader to compare the new information with previous information that is no longer available in working memory.[51] Local coherence is important in order to map elements of the syuzhet as they are being perceived, while global information is important in order to construct the fabula. In non-literary texts, at least two levels of representation of a text also operate at the same time, but are called text-based representation and situation model.

---

[49] Haviland S.E. & Clark H.H., "What's New? Acquiring New Information as a Process in Comprehension," Journal of Verbal Learning and Verbal Behavior, 1974, 13, 512-521. See also Paul van den Broek, "The Causal Inference Maker: Towards a Process Model of Inference Generation in Text Comprehension." In D.A. Balota, G.B. Flores d'Arcais, and K. Rayner (eds.) Comprehension Processes in Reading, Hillsdale, NY: Laurence Erlbaum Associates, 1990, p.436.

> A text-based representation is a representation of the text itself, whereas a situation model is a representation of what the text is about, containing textual information and general world knowledge.[52]

In artistic texts, the situation model gives the global structure for the comprehension of the fabula. Cognitive psychologists proposed that a situation model is established by processing the text into causal chains of inferences, which result in a perception that the text is coherent.[53] In order to determine whether two events in the text exist in causal relations, Paul van den Broek proposes four criteria:

> According to the criterion of *temporal priority*, a cause never occurs after the consequence. According to the criterion of *operativity*, a

---

[50] For a fuller discussion of short term memory capacity limits see George A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," in The Psychological Review, 1956: 63, pp. 81-97.

[51] E.J. O'Brien, M.L. Rizzella, J.E. Albrecht & J.G. Halleran, "Updating a Situation Model: A Memory Based View," Journal of Experimental Psychology: Learning, Memory and Cognition, 1998, Vol. 24, 5, Pp. 1200-1210.

[52] Ibid. P.1200.

[53] Paul van den Broek, op. Cit. P. 423.

> cause is active when the consequence occurs. The *necessity in the circumstances* criterion reflects the fact that if the cause had not happened then the consequence would not have taken place, given the circumstances of the story. The *sufficiency in circumstances* criterion indicates that if the cause occurs, then the consequence will likely occur as well, given the circumstances of the story.[54]

Van den Broek cites much research that supports the existence of these criteria in postulating causal relations. Subjects seem to remember causally related events better than two unrelated events even when the two are adjacent in the surface structure (the syuzhet); in addition, dead-end events (ones that do not move the plot forward, and are not causally related to others) were forgotten faster than causally related events. Summing up much research from the 1980s van den Broek claims:

> Highly connected events are more often included in summaries [. . .], rated as more important [. . .], and retrieved more quickly [. . .], than events with few causal connections.[55]

But the mental representation of the causal relations proposed, or made available by the text is dependent on variable factors. Short-term memory capacity, for instance, may limit the exploration of all possible connections while reading a focal event. Van den Broek proposes a Causal Inference Maker model (CIM) in which

---

[54] Ibid. pp. 424-425. Emphasis in original.

[55] Ibid. P. 429.

> The causal criteria and the limitations determine the content and the types of inferences made, and hence form the conceptual and procedural *constraints* that operate on the inferential process.[56]

Van den Broek's model, then, is based on the assumption that the search for causal relations (according to the four conceptual criteria), and the procedural limitations on attention and memory, guide the reading process, and result in the situation model of the text. This model is a highly selective one, a model that during perception prioritizes causal relations, and claims that these relations affect the long-term status of the memory of the text. That is, events that do not seem as causally related (or potentially so) at the time of perception will be of a lesser status in long term memory (the situation model of the text), and are less likely to alter the comprehension of the text when new information appears. The CIM postulates that when adjacent events provide necessity and sufficiency they will be connected as an inference. When the text provides a break, or a gap, the reader will search the memory for missing information and if found will reinstate an inference. If the text does not provide the necessary and sufficient events for an inference, the reader will fill in the gap by inferences based on world knowledge, a process van den Broek calls elaboration.[57] The model is very useful for a theory of gap filling, but there are two problems with applying it to gap filling in films.

The first problem with van den Broek's model and its application to film has to do with his sole reliance on language:

---

[56] Ibid. P. 433. Emphasis in original.

> It is important to note that several conditions need to be met in order for the reader to be able to construct a coherent mental representation of a text. [. . .] The function and meaning of words need to be identified and the words need to be combined to form a sentence or *proposition.* It is only after these tasks have been achieved that the reader can come to understand the relations among the individual events portrayed in the sentences.[58]

Van den Broek relies solely on language processing, and describes it as a linear and propositional formula, one that contains and defines the logical and causal relations between events. But cinema communicates also with images that provide multiple and overwhelming amounts of detail at once. This information is processed simultaneously as well as serially (see chapter 2), and is stored in memory both as propositional sets and also as depictions (images). But when visual information is retrieved from memory, it is pulled as holistic. In other words, the basic processes that are necessary for a causal memory representation of texts according to van den Broek, do not always occur when the communication channel is visual. We can recall the discussion of Dretske's work (chapter 2) in which he differentiates between sense perception, and meaningful perception. Dretske doesn't go as far as claiming that meaningful perception requires propositional configuration, but he comes close by claiming that meaningful perception requires computation, categorization, and judgement – all high-order cognitive activities. That is, visual

---

[57] Ibid. P. 432.

perception becomes cognitive for Dretske only once these high-order operations take place, and can be stored in memory only when it is "meaningful." But this high-order restriction on the cognition of images is problematic. Images, as I have shown in chapter 2, are processed in a variety of ways, and stored in memory both as propositional sets and as descriptive images. Moreover, because films overwhelm us with a multiplicity of visual details, parts of the filmic image that may not seem relevant at the time of perception may become crucial for narrative comprehension later on. If these visual elements are not stored in memory because they were not categorized as "meaningful," they will not be available for retrieval from memory at the time needed. The moth, in *Silence of the Lambs,* for instance, seem irrelevant to the plot for much of the film's syuzhet, but when Clarice sees one on the serial killer's kitchen counter, it becomes a crucial narrative cue both for her and for the audience (see my discussion in chapter 3). That is, while perceivers assume that any detail (like the moth) in a thriller may turn out to be important, they cannot relate it in a causal fashion to the plot, hence, according to van den Broek are less likely to remember it. But when we see the moth on the serial killer's kitchen counter, we easily and immediately remember and understand the narrative importance of this cue. Unlike natural languages, which are cognitively processed by using high-order cognitive mechanisms, at least some parts of the visual image can (and often do) bypass the categorization and computation process. It is clear that what is problematic for Dretske's meaningful perception is even more problematic for van den Broek's propositional sets. The articulation of images into causal sets is not impossible, but neither is it as easy and automatic as van den Broek suggests it is for reading. More importantly, propositional sets are not necessary, at least not as a prerequisite, for the initial stage of perception and storage in memory of visual information. By prioritizing causal inferences, van den Broek's model leaves out much visual information that could not fit into propositional sets of necessity and sufficiency. But this visual information is clearly often important for narrative comprehension, and we therefore need a more flexible account of narrative construction, one that can allow non-propositional sets to influence the situation model, or the fabula.

The second problem with van den Broek's model can be found in the restriction he places on long term memory, whereby only inferences that are made within the immediate cognitive constraints of memory capacity are stored and used when a gap occurs. In a series of experiments, O'Brien, Rizzella, Albrecht and Halleran provided subjects with a series of sentences that posed a breakup in the coherence of the text. For instance, subjects read a text in which background information indicated that Mary is an avid vegetarian, but later on a sentence described Mary as ordering a cheeseburger. The information about Mary being a vegetarian no longer exists in working memory, but is pulled and is re-activated in order to attempt and solve the problem of contradictory information.[59] The results show that not only causal information is considered for integration in the global and situation model. Instead, any and all memory-based variables can be re-activated and integrated not at the time of perception, but at the time that they become relevant dramatically. Gaps then activate long term memory in an attempt to bring up all information that could be relevant in order to come up with a coherent inference, or a bridge. O'Brien, Rizzella, Albrecht and Halleran's research shows that information is stored in long-term memory even if at the time of perception it seems to be a dead end information (with no causal chain importance). In other words, the hierarchical nature of memory storage and retrieval suggested by van den Broek is challenged here.[60] It is particularly important to film, since film communicates via multiple channels of information, and as I have shown in chapter 3, these channels do not always tell the same story. The verbal causal relations may be different from the visual ones, and the cues of causality may be misleading altogether. For instance, in *Silence of the Lambs* (appendix #5, see full discussion in chapter 3), an inference is made that the FBI is at the house of the serial killer. As I have shown earlier the inference is made by use of the conventions of parallel editing, the inclusion of a title indicating the FBI are in Calumet City, IL, and the two door-bell rings, which connect the exterior of the house, with its interior. But these causal relations are

---

[59] E.J. O'Brien, M.L. Rizzella, J.E. Albrecht & J.G. Halleran, op.cit. pp. 1200-1210. Some experiments used qualifiers in the background information, like "Mary used to be a vegetarian," or "Mary is generally a vegetarian." These cases too, slowed the reading

---

response and showed that readers were activating long term memory, trying to solve a conflict with the term "vegetarian" which was already stored.

[60] Additionally, R. Elio and F.J. Pelletier showed that when contradicting information was presented and belief revision was activated, it was the conditional sentence that was abandoned, rather than the ground (non-conditional) sentence. That is, the syntactic and logical relations determined belief revision, and not necessarily semantic causality. See Renée Elio & Francis Jeff Pelletier, "Belief Change as Propositional Update," Cognitive Science, Vol. 21 (4) 1997, pp. 419-460.

negated when the serial killer opens the door to reveal Clarice (who is in Ohio), and is reiterated moments later, when the FBI breaks into the empty house in Illinois. A causal inference model which restricts memory would make it very difficult (and slow) to understand how Clarice could be at the door of the serial killer. But if other (non-causal at the time) memories are available, the audience is able to correct the inference much faster and more efficiently. The fact that Clarice just had an idea and was going to try and interview a dress maker, just so that she can collect "support evidence" for the trial, clearly becomes important when we see her at the door of the serial killer. The audience then realizes that what we thought was taking place in Illinois (at the basement of the serial killer), was actually happening in Ohio. The access to all background information in memory, which is suggested by O'Brien, Rizzella, Albrecht and Halleran, also resolves the problem of the reliance on language, which was essential for van den Broek's model. If all long-term memory is accessible, then whether it is retrieved propositionally, or pictorially, it is available for gap filling, and for the re-organization of the fabula, or the situation model.

In the case of a film then, gaps activate both visual and verbal memories, and use both of them in a new integration of material, towards a re-organization of the fabula. We shall now return to the findings of the second chapter in an attempt to understand how filmic gap filling operates in practice.

Filling–in Filmic Gaps

In this section I analyze in detail the inference and hypothesis production (gap filling) of two gaps, in two scenes, one from *Dead Poets Society* and the other from *Dangerous Liasions*. This analysis does not attempt to be conclusive, or exhaustive, but I hope that it shows the merit of performing a cognitive analysis of gap filling. There are, of course many other examples and films that could be analyzed as well, so this section should be read as an example of such an application, and as an invitation to further elaboration on this direction of analysis in the future.

Dead Poets Society (Peter Weir, 1989) is a film that climaxes in a gap. The film focuses on the life of teenage boys in an upscale prep school, and particularly the relationship they develop with a charismatic and inspiring teacher, Jack Keating, (played by Robin Williams). One of the main characters is Neil Perry (played by Robert Sean Leonard), a good student from a modest income home, whose real passion is theatre. Neil excels in school, but is still forbidden from participating in a play by his rigid father. Neil disobeys the father's orders, and is extremely successful as Puck ("A Midsummer Night's Dream"). But the father walks in during the play, and afterwards he takes Neil home and informs him that he is to be enrolled in an army academy, so that he is taught discipline. In the following scene Neil commits suicide. But director Peter Weir never shows the suicide (see appendix #3). Weir shows us a series of shots leading to the suicide, and a series of shots after the suicide, and the suicide itself is only confirmed by the film moments later. The film then presents us with a gap, but supports the suicide hypothesis, or inference, that the audience develops. I shall now examine closely the scene, and explain how the gap is being filled.

The scene starts with a shot of Neil standing by the window, shirtless, lifting the window wide open. Eerie music fades in and plays throughout the scene until the moment of the suicide. Neil reaches out to grab the crown of thorns (part of his costume from the play), and puts it on his head. In the next medium close-up (m.c.u. shot #2) he lowers his hands, and looks down. The camera lingers on Neil. While we know the wreath was a prop in the play, the shot creates an allusion to prototypical images of Christ as well. This metaphor does not necessarily operate in the surface structure of the text, since we can easily connect the shots by inferences based on the idea that the wreath is a symbol of the world of theatre which Neil is about to lose. But once we realize (moments later) that Neil has killed himself, the allusion to Christ and the idea of martyrdom emerge quite naturally. Moreover, I believe that even if we don't consciously think of the Christ metaphor at the time of viewing, it still prepares us for the suicide in subliminal ways. That is, the image of Neil standing at the window with the wreath on his head is stored in memory not just as a section in a causal chain that articulates his lamentation over his theatrical career, but also as (at the very least a potential for) a metaphor about martyrdom. The image does not necessarily foreshadow the suicide, but is readily available to support the suicide hypothesis, something that would be unlikely in van den Broek's model.

The scene continues with a series of close-up shots (3-9) on door-knob moving, door opening, feet on the floor, key, and a drawer opening, hands retrieving a wrapped object, and a pan shot from a dresser to a close-up shot of the father sleeping. All these shots are filmed in the dark of the night, and are obscured by the lack of light and the fact that many of them are close-up shots. While the objects in the shots can be identified, it is unclear which space we are in at any given moment, whose feet and hands we see, and what is being retrieved from the locked desk drawer. That is, while the image processing at the ventral stream (object recognition) is more or less readily available, the dorsal stream activity (placing these objects in relations to one another) is much more difficult to pin point, and the mapping of space, action, and who carries the action is at best obscure or ambiguous. Moreover, the gun is never seen, as it is wrapped in cloth. It features in one shot of this sequence (shot 9), when it is being pulled out of the drawer, and the camera zooms out to reveal Neil sitting at a desk holding the wrapped object in the dark. There is no clue at this point that the object is a gun, and there are no visual features that can enable us to come up with recognition (a deconstruction into geons) or a hypothesis that this is a gun. This sequence of seven shots does not lend itself to a connectivity that can inform an inference. The shots seem somewhat dis-connected (both spatially and dramatically), so rather than moving the narrative forward, and together with the eerie music, they result in a sense of vague anticipation of some dramatic event, rather than a clear observation that one is taking place.

In the next shot (#10) the father abruptly wakes up as if from a bad dream, and the music stops. He murmurs something about a sound, but given that we haven't heard anything, we first attribute his concern to a bad dream. Here we assume that we have better means to make an inference than he does. We have not heard anything, and given that he was asleep, we assume that we are cognitively better informed (after all, we are awake. . .). But the father is haunted, and he proceeds to move through the house, turning on the lights everywhere he goes. His search (shots 11-22) is done in full light, and unlike Neil's fragmented and dark journey, the father's is shot mostly in wide-angle shots. We have plenty of visual information: Neil's room is empty, the wreath still on the open window sill, hallways empty, and finally the office is clear and quiet, but the father smells something (shot #18). While the previous sequence was full of ambiguous and difficult to discern information, the father's sequence is visually available, but void of the narrative conclusion we are looking for. Finally, in shot 20, the father (now standing in the office) starts moving to his right, the camera cuts to his point of view shot (#21), which in a pan movement reveals the gun on the floor. The camera keeps panning right, as if the father is moving to the right, and an arm is seen on the floor. In shot #22 the father jumps forward, but the camera is shooting in slow motion, so his movement towards the desk is stretched in time. We never see Neil on the floor, and we are only informed of his death in the following scene, in which the kids at the dorm are waking each other up with the news that Neil is dead. Still, the word suicide is mentioned only much later in the film, when Keaton (the charismatic teacher), is about to be fired.

The gap (the suicide) occurs between shot 9b (Neil sitting at the desk silhouetted) and shot 10a, when the father wakes up abruptly. The filling in of the gap occurs sometime between shot 10, and shot 21 (in which we see the gun and hand on the floor). By the time shot 21 occurs, we are not surprised, but can (partially) confirm the suspicion that Neil shot himself. But given that this information (suicide hypothesis) is never provided in the surface structure of the text, our hypothesis originates from a complex, mostly high-order operation on information available.

Van-den Broek suggests that when we process a text we try to connect adjacent events to one another. As I have shown above, the shots in the sequence leading to the gap do not connect well with one another, and the gap presents a real rupture, a break in the narrative flow. If necessity and sufficiency conditions (of cause and effect) cannot be met, van den Broek suggests that we search memory for missing information that will enable us to reinstate a causal relation and make a bridging inference. But the scene above does not yield itself to those kinds of explicit causal relations either. Neil has not been suicidal throughout the film, there was no mention that the father owned a gun, and Neil's love of theatre (and his father's anger at that) did not provide the main dramatic conflict of the film. There are no direct inferences that can be made based on a memory search then. Moreover, the facts that Tom forbade Neil to participate in the play, that Neil disobeyed his father's orders, and that he is about to be enrolled in an army school could possibly lead to a hypothesis that Neil is leaving home, and that the wrapped package is saved money. But as I shall show shortly, there are reasons why this hypothesis is less likely to emerge.

Van den Broek suggests that in those cases where previous textual information is unavailable, we retreat to our naïve knowledge of the world, thus engaging in a process that he calls elaboration, a process that brings in extra-textual information in order to create the bridge, or fill in the gap. Literary and film theorists discuss an intermediate stage between the long-term memory of the text, and the general world-knowledge, one that relies on knowledge of the genre, and of dramatic texts altogether. Traditional drama, as already identified by the Greeks, contains a set of dramatic conflicts that are usually resolved, but not before a climax occurs. In the sub-genre of prep-school films, a tragedy often occurs, and it is likely to evolve around the unfulfilled hopes and passions of one of the teenage protagonists of the film.

Based on these generic norms, it is likely that while viewing the *Dead Poets Society*, audiences develop an expectation that something "bad" will happen to one of the protagonists. This vague expectation finds a home in the suicide hypothesis, and guides the process that leads to this choice as the preferred inference (over the "escaped home" hypothesis, for instance). But this account of gap filling still seems unsatisfactory. It at best provides us with a general direction for hypothesis production, but does not explain how over the span of 10 shots (at most) we have come to the conclusion that Neil killed himself. I believe that the specificity of the hypothesis has to do with a strong reliance on bottom-up perception of the filmic information in the scene.

The dramatic information in the scene is conveyed almost entirely by visual means (there are a couple of verbal exchanges between the father and mother, but they are not very informative, that is, the father doesn't tell her what he thinks he heard, and why he is up). This visual information is arranged (both in terms of framing, shooting, and editing) to support the suicide hypothesis. The image of Neil as Puck/Christ sets the atmosphere of the scene. The disjointed series of dark and close shots that follows resists full cognitive processing, and therefore support the sense of looming danger. The fact that we see the gun, but cannot know that it is a gun (since it is concealed) prevents us from connecting all the previous shots in a coherent causal chain, but instead gives the impression that the sequence is likely to culminate in a (yet unknown) climax. The visually ambiguous nature of the sequence prevents us from anticipating what this climax is going to be, but prepares us for its coming, and ensures that we are not surprised, but ready to jump to conclusions. The abrupt end of music and the father's jump punctuate the previous sequence, and as the search sequence begins, the audience realizes (based on generic conventions) that the scene is bound to end in tragedy. It is then that the suicide hypothesis formulates, and when we see the gun and the hand on the floor, it is nearly confirmed (Neil could still be just wounded).

The other major support for this hypothesis comes from the eerie music, which stops at the moment of the suicide, and returns when the father sees the gun. In the first 9 shots then, the music seems to indicate a looming danger, or tragedy. But when the music returns in shot #21, it becomes clear that it stood in for Neil's mental state, his decision to kill himself, or for the suicide itself.

I strongly believe that has the suicide hypothesis been solely a result of high-order cognitive activities -- based mostly on causal sets of previously processed and categorized textual information and world knowledge -- it would have been vaguer, and would not have emerged so smoothly during the father's search sequence. But attentiveness to visual detail, and the ability to re-focus and re-interpret it when the narrative conditions change, is a key to the success of the gap filling practices of this scene. What is required here is not just openness to bottom-up information, but an understanding that visual information is not processed cognitively in the same ways that language is. The concealed object, for instance, is categorized as such: an object that could be a gun, a wallet with money, or any number of other objects. It is assumed that this visual memory will be important narratively, but it is not classified as part of a particular propositional (or even meaningful in Dretske's terminology) object. It is stored as an ambiguous object, like much of the items in the dark shots preceding it, and is assumed that the information could become important and explicit at a later point. Visual re-organization works by being able to retrieve images from memory, images that may have not been classified and categorized at the time of perception, but are being determined at the time of this later cognitive operation, the one of gap filling. The scene is effective (that is, not confusing) and elegant because high-order operations work in tandem with bottom-up perception.

Dangerous Liasions (Stephen Frears, 1988) is a drama (based on the epistolary novel by Choderlos de Laclos) set in 18th century France. The film tells the story of two aristocrats and ex-lovers, the Marquise De Merteuil and the Vicomte De Valmont (played by Glenn Close and John Malkovitch respectively). The two have an elaborate scheme of sexual pursuits and seductions (set in a sort of competition between the two of them) that is supposed to eventually lead them to have sex with one another once more. The film keeps the audience in an epistemically privileged position with regards to the plans of the two, and therefore at an advantageous perspective with regard to all other characters in the film, particularly the objects of the schemes. In the three scenes discussed below Valmont is trying to seduce a devout Christian, and a married woman, Mme. De Tourvel (played by Michelle Pfeifer). Mme. De Tourvel seems to be quite disgusted with Valmont and his reputation, and the film never lets the audience know what she thinks or feels aside from what we see happening during their encounters. The film then posits a gap with regards to how Mme. De Tourvel feels towards Valmont. In other words, Mme. De Tourvel does not reveal her cards, and the audience cannot deduce a hypothesis about her future plans, or about her present emotions, by relying on the textual information provided by the film. Using generic conventions, the audience may assume that in such a drama characters are doomed to fall in love with one another, but Mme. De Tourvel's actions and behavior do not support or negate this expectation.

The first scene discussed (appendix #4, seduction#1) focuses on the daily walk the Madame takes through the gardens of Valmont's aunt. Against her will, Valmont joins her on the walk and they have a short argument about who incriminated his name in her eyes. The scene consists of a short introductory shot (#1), and a lengthy shot (#2) in which the two of them are walking on a path, the camera tracking back ahead of them. The shot design keeps the flustered Mme. De Tourvel in the foreground, while Valmont is changing from her right to the left side, six or seven times, walking slightly behind her. Mme. De Tourvel walks straight ahead, and refuses to look back at him. She is centered, tries to stay focused, and her body language conveys a sense of self-purpose and dignity. Valmont, on the other hand, seems mentally erratic (through his dialogue), and visually unstable (jumping from right frame to left frame continually), always trying to catch up with Mme. De Tourvel.

In a later scene in the film (Appendix #4, seduction #2) the two are seen again walking together. The camera (shots 1, 3) is in the same position as in the previous scene, tracking back while the two are walking towards it. The scene is framed with a voice-over of Valmont, reading a letter he wrote to the Marquise De Merteuil (who is seen reading it) about his progress with Madame De Tourvel. The voice over indicates that Mme. De Tourvel has accepted his love, and that they go on a daily walk on "the path of no return." The camera cuts to the walk, and Valmont is flattering Mme. De Tourvel, telling her he has changed significantly under her Christian influence. The audience knows he is lying, as we see him spending his time seducing a teenage girl, or going back to Paris, for *rendez-vous* with sexual partners. Mme. De Tourvel laughs when Valmont says that he is "more celibate than a monk," but it is unclear whether she believes him or not. There is still no indication as to her emotional state, not verbally or by her body language. But the visual structure of the scene suggests that the seduction plan is advancing. In this scene, Mme. de Tourvel and Valmont are walking side by side, thus he gains visual power, stability, respectability, and more importantly, he is presented on the same visual plane as her. This subtle visual structure conveys the message that the affair is on its way, and fills in a narrative gap as to Mme. De Tourvel's feelings. When the voice over of Valmont returns at the end of the scene saying "I feel she is inches from surrender," the audience is not surprised, but prepared.

The sequence culminates a little while later with an indoor scene in which Mme. De Tourvel admits that she has fallen in love with Valmont. In the scene Valmont finds a rare opportunity to speak with the Madame alone, and he uses it to push her to a break-down. But even before she admits her love, the visual structure conveys it to us. To start with, Mme. De Tourvel is not as self-assured anymore, and while her sense of purpose and direction was firm on the first two walks, she now struggles to gain her strength, turning away from Valmont aimlessly (shot 1b, 1c, 1d, 1g, 1h). Her unfocused movement around the frame and around Valmont is compounded by his move around her. From the beginning of shot 1, Valmont places himself behind her, and he follows her around until 1e. In 1f Valmont starts circling around the Madame, and she around him. The dizzying movement is further emphasized by the camera movement, which pans with Mme. De Tourvel to the right and to the left of the room. The camera then, which mirrored Mme. de Tourvel's sense of direction and dedication in the symmetric shots in the garden, is now mirroring her crumbling mental and emotional state. Throughout the first shot, the camera and Valmont manage to visually trap Mme. de Tourvel between them (he usually slightly behind, the camera blocking her movement forward), and by the end of the shot she looks like a trapped animal, a prey. The scene concludes with a series of 6 over the shoulder shots, and in shot number 5, she finally looks at his eyes and says "yes," thus admitting she loves him. But her admission seems redundant, as the audience already knows Mme. De Tourvel loves Valmont, because the visual narration has already told us so.

The gap in *Dangerous Liasions* is very different than the gap in *Dead Poets Society*. While the latter film never shows the suicide, the former actually provides all the information needed in order to make inferences about Mme. De Tourvel's feelings. But this information is provided visually, and even within the visual track it is done not by overt body language or action, but by subtle metaphorical relationships between camera-work and character. I therefore call this kind of gap an illusory gap, one that provides the necessary information to make inferences, but does not make us consciously aware that it does so. In the following paragraphs, with the aid of the cognitive material on visual and verbal perception, I explain why the illusory gaps seem real, or pass as real ones.

As I described in the second chapter, natural languages are based on a lexicon of words, which stand in for objects or concepts in the world.

The signs (words) which represent signifieds are arbitrary and abstract, and have been coded to signify concepts as a result of a long social and historical process. Natural language perception therefore requires high-order mechanisms such as categorization, matching to previously stored items, and the decoding and encoding of phonemes, morphemes, and syntax, to form or understand meaningful (semantic) messages. Philosophers, objectivist linguists and cognitive linguists argue about the nature of the relationship between the structural aspects of language (phonemes, morphemes and syntax), and semantics. Particularly, much debate has been focused on whether semantics is independent of syntax (the objectivist view) or whether it influences syntax and other structural aspects of language, and particularly of language use. These positions have many consequences on our understanding of perception and cognition of language, because they presuppose different sets of necessary mental operations at work. Either way, though, language is delivered linearly, and it is based on propositional-like sets of subject, object, action, etc. Cognitive perception of language then also necessitates the use of high-order cognitive mechanisms that translates language into these propositional sets. These operations also make language comprehension more consciously available to the perceiver. That is, given that words need to be decoded, then matched to lexicon items in memory, and content of sentences has to be classified in logical cause and effect relations, the awareness to the perceptual processes is relatively high. As an example, if a person hears a word she does not know, it is clear to that person that this word was not learned and does not exist in their lexicon.

One then may attempt to understand the word by inferring from the context (realizing it is a noun, or a verb, etc.), or one may learn the content by a metaphorical process (through a dictionary or asking another person). It is clear though that until one knows what the sign stands for, comprehension is likely to be hampered.

Visual perception and cognition, on the other hand, operates in quite different ways. Images are not learned in the same way that language is. As Dretske has claimed, we have the ability to have meaningful perception -- discern objects, movement, spatial relations, etc. -- without needing to know what these things are by name. That is, seeing may not necessarily require propositional configuration of the material perceived in order to be meaningful. Upon seeing an image we are overwhelmed by a multiplicity of information that needs to be processed. As I have shown in chapter 2, some of it is processed linearly (for instance by looking for patterns), but some aspects of the image are processed holistically. The visual field is also processed in two different parallel tracks, the ventral and the dorsal streams. The ventral stream is responsible for object recognition, while the dorsal stream deals with spatial relations. Object recognition is done mostly by some abstraction into prototypes, or basic shapes like geons. But the relations between the objects in the visual field are complex. A large visual environment is mapped by at least two visual searches: one focuses on grouping objects into distinct groups that are more easily and speedily processed; but at the same time, the visual field is scanned for features that need individual processing. In a film, a space is usually categorized after a quick serial scan (a room in Paris, a train station in the West), and the focus is turned towards action, objects that dramatically stand out, and clues for further narrative development. While action in film can easily be summarized in propositional sets (who did what to whom), and it lends itself to a linear reading, as it is set in a temporal axis, the rest of the visual properties do not necessarily lend themselves to these linear, cause-and-effect procedures. The perception of images then requires a set of very different cognitive capacities, and much of it operates from the bottom-up, rather than from the top-down. Images are not abstracted and conceptualized in order to be perceived and comprehended. They are seen "as is" and processed as such. But this bottom-up seeing, of holistic, rather than propositional networks, has an effect on the level of awareness we have of visual perception. Unlike language, which is learnt and codified, the perception of images seem to function in automatic and unconscious ways, and as viewers we are therefore less aware of its cognitive processing for higher order operations.

In addition to the differences in perception, some differences have also been noted in the memory storage mechanisms of visual and verbal information. While the verbal information is already codified and abstracted, and therefore can quite smoothly be categorized in memory for storage, there is a big debate about the nature of visual memories. But whether they are propositional or depictive while stored in memory, we now have enough evidence that when retrieved, visual memories are used as descriptions and not as propositional sets. That is, while we retrieve a visual memory, we operate on it with the same mechanisms as at the time of perception, namely those that use attention to both holistic and analog information. At the time of retrieval then, we are again relying on bottom-up processes, more than on top-down beliefs, and well formulated structures (like syntax, lexicon, logic, etc.). This data leads to the conclusion that we are less consciously aware of visual processing than we are of verbal messages. Consciously aware or not, however, the visual cues are cognitively processed and affect our judgements about the events that happened and/or are about to happen.

While watching the seduction scenes in *Dangerous Liasions* we receive and process the visual cues that lead us to believe that Mme. De Tourvel is falling in love with Valmont. The visual cues (those of body language, camera work, etc.) are processed from the bottom-up, and result in the high-order belief that Mme. de Tourvel has fallen in love with Valmont. But because of the unaware nature of visual perception, we are unable to pinpoint why we come up with this particular inference, and we therefore assume that there was an informational gap. But as I have shown above, a gap did not exist, and hence the term, illusory gaps.

Summary

In this chapter I have shown that reception theories focus on either how the text constitutes the reading process (bottom-up approach), or on how the perceiver constitutes the text (top-down operations). But through numerous examples I have shown that perception and interpretation are not distinct and separate activities, but are inter-dependent on one another. As an example of comprehension and interpretation practices I focused on the phenomenon of gap filling.

The analysis of two scenes that employ gaps (from *Dead Poets Society* and *Dangerous Liasions)* shows that the activity of the perceiver is complex and depends both on bottom-up perception of visual and auditory information, and on top-down assumptions about the genre, and the nature of drama in general. While interpretation requires high-order cognitive operations -- such as re-organization of story data in cause and effect structures, retrieval from memory of past events, speculation, and hypotheses production -- it nevertheless operates in tandem with bottom-up perceptions, and is very sensitive, and very ready to adjust to the new flow of textual information. In addition, I have shown that the channel of information (i.e., visual or verbal) affects the perception and cognition of that information, and determines whether we categorize a dramatic ambiguity as a gap. Given that we are less consciously aware of visual information processing, we may assume that a gap exists, while the dramatic information is actually presented, only via a less aware channel of information. I call this case illusory gaps, one that is assumed by top-down assumptions, but is not present at the level of the cinematic style. Whether illusory or real, bridging the gaps relies on the information available to memory, which in turn, relies on the specific channel of perception. Gaps are therefore a good example of how bottom-up perceptions interact with top-down assumptions to yield interpretations.

Conclusion

In this conclusion I will first give a general summary of the thesis, then a short description of the particular claims made in each chapter, and finally I will briefly discuss some of the possible applications of such claims, or suggestions of extensions of this research to future projects.

Thesis

Many film and literary theorists have tried to articulate the relationship between the perceiver and the text. The following quote by Francesco Casetti is one typical example:

> By connecting sparse cues in order to build up a character or place; by providing a framework in which the data are given their full value (e.g. suggestiveness of genre labels); by scanning visual patterns within the frame to grasp the essential and discard the unimportant; by filling the gaps in the narrative to restore the completeness of the story (frequently the unseen is essential in explaining what seems manifest). The viewer lives in the film [. . .] The spectator *commits* him/herself to viewing.[1]

But Casetti cannot explain why or how the spectator lives in the film, or what exactly is the meaning of committing to the viewing experience. Whether looking at the interaction between the perceiver and text as dominated by the perceiver's activity (as Casetti does), or as cued primarily, or solely, by the text (as the structuralists do), most theorists do not give comprehensive accounts of the nature of this interaction.

---

[1] Francesco Casetti, "Looking for the Spectator," Iris 1,2, (1983), 24-25.

In this dissertation I have claimed that cognitive science research -- particularly research on the issues of perception of textual information, and of high order cognitive activities such as the organization of the narrative (much in Casetti's spirit) -- can shed light on the matter. In particular, I have showed that both in the organization of textual material (narration), and in the decoding of story events (comprehension), bottom-up perceptions and top-down expectations interact in complex, and inter-dependent ways.

I have also made a second claim, one that has also emerged out of research in cognitive science. The claim is that differences in the perception and cognition of images and of natural languages are significant for many cognitive operations performed on that information, and thus for film comprehension. Language is abstract, highly coded, processed serially, and stored in memory both in categories and in propositional sets. Images, on the other hand, are processed both holistically and serially, are probably abstracted for storage in memory, but are used as depictive representations at the time of high order cognitive mental operations. These cognitive differences may seem sensible and reasonable if one thinks of the many differences between language and images. But the differences in the perception and cognition of language and images pose a problem to film theory. Structuralism, semiotics and film narratology have all emerged out of literary and linguistic models, and have thus struggled to address the cinematic image in its own right. As a result, a full account of film's (linguistic, musical, and image,) perception has not been proposed. Cognitive film theorists have done much to address the problems of semiotic and linguistic based theories, but have, as a result, treated cinema as a visual medium, without addressing its linguistic attributes. In this thesis I have attempted to account for the interaction of images and natural languages, and how together these two different tracks of information narrate film. In particular, I have tried to account for some of the impact of the material choice of narrative information (i.e., image or language) to inference making, hypotheses production, and other activities associated with comprehension. I believe that a cognitive understanding of film's sense experience and of narrative comprehension greatly enhances our understanding of the filmic medium, and sheds light on areas of perception and spectatorship that have not been explored before. In the following pages I shall briefly recount some of the more specific conclusions of this thesis, conclusions which support the two claims I made above.

Chapter Summary

In the first chapter I demonstrated the problem with film narratology, namely, that it relies too heavily on literary narratology, and therefore tends to ignore or simplify the discussion on the actual material filmic presentation, that is, images and sounds. Through a discussion of the concepts of focalization, enunciation and point of view, I have shown that what might apply for literature does not hold in the case of film.

In the second chapter I reviewed literature from cognitive sciences, in an attempt to portray the differences in the perception, cognitive processing, and memory storage of images and language. To sum up some of the conclusions from this chapter: natural languages are based on a lexicon of words, which stand in for objects or concepts in the world. The signs (words) which represent signifieds are arbitrary and abstract, and have been coded to signify concepts as a result of a long social and historical process. Natural language perception therefore requires high order mechanisms such as categorization, matching to previously stored items, and the decoding and encoding of phonemes, morphemes, and syntax, to form or understand meaningful (semantic) messages. Language is delivered linearly, and it is based on propositional-like sets of subject, object, action, etc. Cognitive perception of language then also necessitates the use of high-order cognitive mechanisms that translate language into these propositional sets. These operations also make language comprehension more consciously available to the perceiver. That is, given that words need to be decoded, then matched to lexicon items in memory, and content of sentences has to be classified in logical cause and effect relations, the awareness to the perceptual processes is relatively high.

Visual perception and cognition, on the other hand, operate in quite different ways. Images are not learned in the same way that language is. As Dretske has claimed, we have the ability to have meaningful perception -- discern objects, movement, spatial relations, etc. -- without needing to know what these things are by name. That is, seeing may not necessarily require propositional configuration of the material perceived in order to be meaningful. Upon seeing an image we are overwhelmed by a multiplicity of information that needs to be processed. As I have shown in chapter 2, some of it is processed linearly (for instance by looking for patterns), but some aspects of the image are processed holistically. While action in film can easily be summarized in propositional sets (who did what to whom), and it lends itself to a linear reading, as it is set in a temporal axis, the rest of the visual properties do not necessarily lend themselves to these linear, cause-and-effect procedures. The perception of images then requires a set of very different cognitive capacities, and much of it operates from the bottom-up, rather than from the top-down. But this bottom-up seeing, of holistic, rather than propositional networks, has an effect on the level of awareness we have of visual perception. Unlike language, which is learnt and codified, the perception of images seem to function in automatic and unconscious ways, and as viewers we are therefore less aware of its cognitive processing for higher order operations.

In addition to the differences in perception, some differences have also been noted in the memory storage mechanisms of visual and verbal information. While the verbal information is already codified and abstracted, and therefore can quite smoothly be categorized in memory for storage, there is a big debate about the nature of visual memories. But whether they are propositional or depictive while stored in memory, we now have enough evidence that when retrieved, visual memories are used as descriptions and not as propositional sets. That is, while we retrieve a visual memory, we operate on it with the same mechanisms as at the time of perception, namely those that use attention to both holistic and analog information. At the time of retrieval then, we are again relying on bottom-up processes, more than on top down beliefs, and well formulated structures (like syntax, lexicon, logic, etc.). This data leads to the conclusion that we are less consciously aware of visual processing than we are of verbal messages. Consciously aware or not, however, the visual cues are cognitively processed and affect our judgements about the events that happened and/or are about to happen.

With the insight about the differences between the cognitive processing of images and natural language, I set out in the third chapter to propose a cognitive extension to structural models of film narration. First, I showed some fundamental problems with film semiotics, and particularly with its articulation by Christian Metz. Semiotics emerged as a field related to linguistics at the turn of the century. Semioticians attempted to explain all sign systems (linguistic, visual, sign language, body-language, etc.,) as based on the same structural elements. A sign is comprised of a signifier (a representation), a signified (an idea being communicated), and a conventional practice that binds them together. And while it is easy to see this system at work with natural languages, it is harder to apply it to images. Images are not conventionalized like language, and they do not require for their deciphering the knowledge of a lexicon. Using many examples from cognitive film theorists such as Anderson (ecological approach), Noël Carroll, and Gregory Currie, I have shown that the cinematic image does not behave like the semiotic sign. Film images are understood in much the same way that our environment is understood, which is natural, rather than conventional, and bears more ties to biological developments and survival needs, than to cultural ones. Semiotics then, can shed some light on film, but cannot account for the whole cinematic experience, or for the fact that images are not good representatives of the semiotic sign.

The semiotic model is tightly related to structuralist narratological models. While not necessarily linguistic-based, structuralist film theory posits complex architectural structures to describe artistic texts. The structuralist project aims to understand how style (images and sounds) are formed together to present the *syuzhet* (story events as they unfold in the text in linear order), and how the *syuzhet*, in turn, is transformed in the mind of the perceiver to convey the *fabula*, or the plot of the film. These structures can be understood only by high-order cognitive operations, ones that operate on cause-and-effect structures, re-arrange temporal events, and come up with logical connections. It is clear that the *fabula* is a high-order cognitive construct, since it requires memory, problem solving techniques, and other complex cognitive operations for its construction. But structural narratologists tend to focus on those high order operations, and avoid a discussion of how these operations are influenced by bottom-up perceptions. David Bordwell, for instance, proposes a narrative model that he calls "the bull's eye schema." The model centers on the characters and their activities in the diegetic world, and then expands out to account for visual spaces, camera work, extra-diegetic music, etc. This model tries to cohere the narrative around the characters' presence and actions, and as I have shown, it is not a very good model to describe films like *Breathless* or *Rashomon*. This is because Bordwell prioritizes highly coded narratorial information, and ignores low-level perceptions, particularly ones that do not support the "bull's eye schema."

Edward Branigan proposes a much more comprehensive structural model of narration. This model is based on epistemic levels (how does the perceiver know what she knows), and a system of embedding of these levels. But this account does not discuss bottom-up perceptions either. When discussing point of view, Branigan rejects any romantic notions of character's focalization, and he describes this device as a textual tool, designed to provide or restrict the knowledge of the perceiver. But because Branigan does not take into account the cognitive aspects of point of view editing that is, that it imitate innate biologically -- and very important for survival -- already existing looking patterns. The emotional and ideological identification with characters and states of affairs cannot be explained solely by an appeal to the epistemic status of the perceiver (a very high-order cognitive activity), but needs to take into account lower-level perceptual and cognitive reactions as well.

Literary narratology is full of examples of unreliable narrations, a term which usually refers to narrators whose values and claims are undermined by the implied author. But in film, unreliability is complex, and more often than not, the implied author is not as present as in literature. Gregory Currie discusses unreliability and ambiguity, and as I have shown, a cognitive account can explain why filmic ambiguity is easier to achieve than negation of narrator's reliability. Images are perceived as is, and because they have no syntax or grammar, they cannot negate in the same way that language does. For visual film narration to negate, it will have to employ another mode of narration like dialogue, or the use of editing conventions. In ambiguous narration, on the other hand, the text lends itself to several interpretations, and it refrains from giving the perceiver enough information to decide which interpretation is preferred. A film can then present a few visual versions, and leave their epistemic status open, leading the perceiver to an ambiguous interpretation.

In the rest of the chapter I discussed focalization, narration and point of view from a cognitive perspective. Through a detailed analysis of two scenes from *Rambling Rose* I extended Branigan's model of narration to be sensitive to the actual channel of information, and to the interaction between high-order cognitive activities and bottom-up sensory perceptions.

In the fourth chapter I focused on the activities the perceiver engages in while constructing the narrative of a film. In particular I looked at the production of hypotheses when the text presents the perceiver with an informational gap. I first reviewed the existing positions in reception studies, namely the text-oriented approach (such as Iser, Ingarden and Jauss), and the historical materialist approach (that of Fish and Staiger). Textual-oriented reception studies (such as Iser's) differ from the historical approaches (such as Staiger's or Fish's) in the following ways: while textual approaches look at an object (text), the historical approach looks at an event (perception); while the former looks for meaning, the latter examines functions; and finally, while the textual group constructs a universal reader, the other group looks at an actual historical one. We can sum it up by saying that for Iser, the text and its perception determine the interpretive process, while for Staiger, ideology determines perception and interpretation. In cognitive terminology, we can posit this debate as one between top down emphasis (beliefs determine perception) to bottom-up approach (textual information leads to higher cognitive activity such as interpretation). But whether these researchers support a top-down or a bottom-up approach, they refrain from articulating a specific relationship between the two operations. In this chapter I have shown that while the fabula and its construction require many high-order cognitive activities, these are not independent from actual consumption of textual information. As an example, I focused on the issue of gap filling, whereby the perceiver is required to hypothesize as to what has happened, in order to cohere the text. While hypotheses production relies on top-down beliefs, like world knowledge, or generic expectations, it is also very sensitive to textual information, and is always in flux. Belief revision is common, and the perceiver is ready to alter its hypotheses when new information contradicts previous assumptions. The structural models explain gap filling mostly as a high order cognitive activity, but that account cannot explain the speed and ease with which we change our beliefs. In discussions of *The Silence of the Lambs, The Usual Suspects,* and *Breathless* I exemplify these problems.

Gap filling naturally relies on high-level cognitive operations, whereby plot events are ranked according to their relationships, and are re-organized in cause-and-effect structures. And while action and language are linear and lend themselves to be coded in such propositional sets, images, as the second chapter shows may resist this categorization. Using research from cognitive psychology, and scene analyses of various films, I have shown that during the activity of gap filling, visual and verbal memories are pulled and used in their respective formats, and whether they have been coded in high-order cognitive mechanisms or not, they are available for hypotheses production. In fact, the analysis of the suicide scene from *Dead Poets Society* shows that bottom-up perception is very much active during the filling-in of the gap. Moreover, because most of the clues as to how to fill in the gap are visual and are provided in obscure lighting and viewing circumstances (night, darkness, framing), these clues are not coded until the hypotheses emerges, and all the details fit together in it nicely. That is, the image track is processed and remembered in its ambiguity, and it receives full narrative clarity only once the hypothesis of suicide is formed. Visual re-organization works by being able to retrieve images from memory, images that may have not been classified and categorized at the time of perception, but are being determined as meaningful at the time of this later cognitive operation, the one of gap filling. The suicide scene is effective (that is, not confusing) and elegant because high-order operations work in tandem with bottom-up perception.

In addition I showed that in some cases, what seems to be narrative gaps are not gaps at all. The perceiver is under the impression that narrative information is missing, but this information is provided in the visual track, and the perceiver is just less aware of its presence. An analysis of scenes from *Dangerous Liasions* shows a case of such an illusory gap, and explains how although all the information is provided, the impression of a gap exists. I am confident that if a cognitive approach to film music is incorporated to such a project, other classes of illusory gaps would be explained in similar terms.

In this chapter I formed an intuition about why sound effects are often remembered as visual information. While much more work needs to be done on sound effects, it seems that certain sound effects serve as devices that expand the boundaries of the frame, indicating events, or spatial cues that are not provided visually. These sound effects are processed much like invited inferences (inferences that are made by the perceiver, but are remembered not as hypotheses, but as information that was actually provided). And because these sound effects' primary role was to expand the visual frame, they are remembered as visual spaces.

Future Directions

It is clear at this point that a complete cognitive account of film narration and comprehension needs to include a thorough description of music and sound effects processing as well. It is my hope that I can expand the work that I have done here to include these channels of information, as well as further develop the work already done on visuals and language. Once we have an understanding of the perception and cognition of all the material aspects of film, we can propose a complete model of narration, one that accounts for the specific channel of information. This model will address methods of memory storage, and other specific mechanisms that are at work while we process music, sound effects, language, and images. In addition, this model will enable us to understand perceivers' levels of awareness to the information delivered (verbal, visual, musical, etc.), that is, how conscious they are of perceiving this information and processing it. The level of consciousness has an effect on the awareness to the process of inference making, and thus on awareness to how we come up with the narrative of the text. The model can answer the same question Branigan asks: *what* does the perceiver know, and *how*. But the answer will not rely only on epistemic status, but also on the specificities of processing each channel of information.

A cognitive account of the film experience can also resolve some of the debates between the textual branch of "reader-oriented" research, to the materialist "reception-studies" one, as it will define more accurately the relationship between a top-down approach and a bottom-up one. Once we have a broad understanding of all tracks of information, we can perform detailed analysis of scenes that tend to yield themselves to multiple interpretations. These analyses will outline the narratorial devices used, and will be able to locate moments of ambiguity, contradictory messages, or gaps. These moments naturally invite an interaction between top-down procedures, such as belief and expectation, and bottom-up perception. Such a localized reading of moments of indeterminacy will at the very least bring the debate into concrete grounds, if not resolve it altogether, as each instance would determine different relations between top-down and bottom-up operations.

A full cognitive model of narration will hopefully also help to explain why a film's ending creates closure in a uniformed way. My intuition is that the text's ending has very few gaps, and is mostly designed to fill in previous gaps, that is, to confirm or cancel hypotheses that were already created. The different sensory tracks all combine at film's end to tell the same plot information. Explicit textual information with no gaps encourages bottom-up perception, which tends to lead to a similar comprehension by different perceivers.

Using this model, I hope to further explore issues of genre and expectations and how "art" cinema is perceived in comparison to generic films. Hypotheses production in the case of a genre film is guided by a strong top-down process of generic expectations. The audience's knowledge of the general structure of the plot, themes and set of characters, turns the viewing attention to subtleties of plot or character developments. The role of gaps and hypotheses, then, is not so much in figuring out where the film's plot is heading, but in noticing the changes and developments in the genre as a whole. In "art" cinema, on the other hand, every viewing experience is focused on the particular plot structures the film employs, and the role of hypotheses production is crucial to the understanding of the narrative. Moreover, since "art" cinema often avoids dramatic or discursive closure, the text's comprehension is left more open to individual interpretations. This "openness" of the text can now be explained through textual structures (which influence perception), rather than through reading practices, and is therefore a good example for the benefits of a cognitive model of film comprehension.

On a more global level of film theory, the insight we can obtain from a cognitive account of film narration and comprehension is, I believe, enormous. A full cognitive account of narration will explain some texts in grounded and simpler terms than exist today. My brief discussion of feminism, while analyzing *Rambling Rose* was aimed at exposing that kind of a direction. While a feminist approach to the film may find many ideological biases, a cognitive approach explains some of these biases in structural-narratorial choices, rather than in conspiratory ideological (in this case patriarchal) ones. In other words, a feminist (or for that matter post-colonial, queer, etc.,) reading of film concentrates on stereotypes, character and audience positioning with regards to one another, etc. A cognitive account, on the other hand, looks at how information is delivered (the track of information), and also how the perceiver knows what she knows. A cognitive account of film narration and comprehension then grounds the ideological discussion in actual understanding of the perception of material filmic information, and in the epistemic status of the perceiver.

But a cognitive account of comprehension and other activities the viewer is engaged in can also shed light on other psychological and cognitive operations that are active during the viewing time. The only sustainable account of viewing, up to cognitive film theory, is that of psychoanalysis, which is a partial and problematic account to begin with. The psychoanalytic account is based on regression to pre-ego stages of development of subjectivity (the mirror phase in Lacan's articulation), and is at best speculative, and cannot be clinically proven. It also sets the audience in relations of desire to the alter-ego presented on the screen, and the focus on desire limits the scope of other mental and emotional processes at work during viewing. More work in cognitive film theory can give us a grounded understanding of viewing with its cognitive, emotional, and ideological aspects taken into account as a whole.

Once a full model of film perception and comprehension is available, it will be possible to conduct experiments to see how the channels of information interact, and how the levels of awareness to the cognitive processing of different channels affect inference making and other high-order cognitive activities. Experiments can focus on showing the same filmic story in different versions, each using different channels of information (i.e., with or without music, certain things said, others not, etc.). Alternatively, experiments can focus on inference making, and verbal reports of subjects on what led them to make these inferences. Hypotheses changes (such as in the case of *The Usual Suspects*) can be studied as well, followed by verbal accounts of subjects, and scene analysis in the form I have performed here.

I hope that in this thesis I have shown some of the potential to such an approach to film theory, and that with more time and work devoted to the issues outlined above we will have a better understanding of cinema.

Bibliography

Akmajian, Adrian & Richard A. Demers, Ann K. Farmer and Robert M. Harnish, Linguistics: An Introduction to Language and Communication, Cambridge, MA: MIT Press, 3rd Ed. 1990.

Allen, Richard Projecting Illusion: Film Spectatorship and the Impression of Reality, New York, NY: Cambridge University Press, 1995.

Allen, Robert, "Reader Oriented Criticism and Television," in (ed.) Robert Allen, Channels of Discourse: Television and Contemporary Criticism, Chappel Hill: University of North Carolina Press, 1987.

Anderson, Joseph D. The Reality of Illusion: An Ecological Approach to Cognitive Film Theory, Carbondale and Edwardsville: Southern Illinois University Press, 1996.

-------------------- & Barbara Anderson, "Motion Perception in Motion Pictures," in (eds.) Theresa de-Lauretis and Stephen Heath, The Cinematic Apparatus, New York: St. Martin's Press, 1980.

Anderson, J. R. "Acquisition of Cognitive Skill" Psychological Review, 1982: 89.

Austin, J.L. How To Do Things With Words, ed. J.O. Urmson and Marina Sbisa, Cambridge: MIT University Press, 1975.

Bal, Mieke Narratology: Introduction to the Theory of Narrative, Toronto: University of Toronto Press, 1985.

Barnouw, Dagmar "Review of *The Act of Reading* and *The Implied Reader* by Wolfgang Iser," Modern Languages Notes 94 (December 1979).

Barthes, Roland "The Death of the Author" in David Lodge (ed.) Modern Criticism and Theory, Longamn, 1988.

---------------------, "Introduction to the Structural Analysis of Narratives" in Image-- Music-- Text, trans. by Stephen Heath, New York: Hill & Wang, 1977.

Baudry, Jean Louis "Ideological effects of the Basic Cinematographic Apparatus," Film Quarterly 18, 2 (winter 1974/75).

Bazin, André What Is Cinema Vol. I & II, Berkeley, University of California Press, 1967.

Bennett, Tony, "Text and Social Process: The Case of James Bond," Screen Education 41 (Winter/Spring 1982).

Benveniste, Emile "Subjectivity in Language" in Problems in General Linguistics, trans. by Mary Elizabeth Meek, University of Miami Press, 1971.

Biederman, Irving "Visual Object Recognition" in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995.

Black, David Allan "Genette and Film: Narrative Level in the Fiction Cinema," Wide Angle, 8 (1986): 19-26.

Booth, Wayne The Rhetoric of Fiction, Chicago: University of Chicago Press, 1961.

Bordwell, David "A Case for Cognitivism," Iris 9 (Spring 1989): 11-40.

---------------------"A Case for Cognitivism: Further Reflections," *Iris* 11(Summer 1990): 107-12.

---------------------. Making Meaning: Inference and Rhetoric in the Interpretation of Cinema, Harvard University Press, 1989.

---------------------. Narration in the Fiction Film, Madison: University of Wisconsin Press, 1985.

------------------- & Noël Carroll, eds. Post Theory: Reconstructing Film Studies, Madison: University of Wisconsin press, 1996.

Bower, Gordon and Randolph K. Cirilo, "Cognitive Psychology and Text Processing," in Handbook of Discourse Analysis 1: Disciplines of Discourse, ed. by Teun A. van Dijk, New York: Academic Press, 1985.

Boyd, David, Film and the Interpretive Process, New York: Peter Lang, 1989.

Branigan, Edward "Here is a Picture of No Revolver; The Negation of Images, and Methods for Analyzing the Structure of Pictorial Statements," Wide Angle 8, 1986.

----------------------, Narrative Comprehension and Film, London & NY: Routledge, 1992.

----------------------. Point of View in the Cineam: A Theory of Narration and Subjectivity in Classical Film, Mouton Press, 1984.

----------------------"Sound and Epistemology in Film," Journal of Aesthetics and Art Criticism 47 (Fall 1989): 311-24.

Brooks, Peter, Reading For The Plot: Design and Intention in Narrative, New York: Knopf, 1984.

Brooks, Virginia, "Film, Perception and Cognitive Psychology," Millenium Film Journal (Fall/Winter 1984-5).

Brown, Royal S. Overtones and Undertones: Reading Film Music, University of California Press, 1984.

Browne, Nick, The Rhetoric of Film Narration, Ann Arbor, MI: Michigan University Press, 1976.

Bruce, Vicki, and Patrick R. Green, Visual Perception: Physiology, Psychology, and Ecology, 2nd ed. Hillside, N.J.: Lawrence Earlbaum, 1990.

Bryant, Jennings, and Dolf Zillmann, eds. Responding to the Screen: Reception and Reaction Processes, Hillside, N.J.: Earlbaum, 1991.

Buckland, Warren, ed. The Film Spectator: From Sign to Mind, Amsterdam University Press, 1995.

Burgoyne, Robert, "The Cinematic Narrator: The Logics and Pragmatics of Impersonal Narration," Journal of Film and Video 42 (Spring 1990).

---------------------, Sandy Flitterman-Lewis, and Robert Stam, New Vocabularies in Film Semiotics: Structuralism, Post-Structuralism and Beyond, London: Routledge, 1992.

Butterworth, George & Lesley Grover, in Noël Carroll, Theorizing the Moving Image, New York, NY: Cambridge University Press, 1996.

Carr, T. H. "Consciousness in Models of Human Information Processing: Primary Memory, Executive Control, and Input Regulation," in G. Underwood & R. Stevens (eds.) Aspects of Consciousness, London: Academic Press, 1979.

------------. "Perceiving Visual Language" in Handbook for Perception and Human Performance II, ed. by Keneth Boff, Lloyd Kaufman, and James P. Thomas, NY: John Wiley and Sons, 1986.

Carroll, Noël, Mystifying Movies: Fads & Fallacies in Contemporary Film Theory, New York: Columbia University Press, 1988.

---------------, Philosophical Problems of Classical Film Theory, Princeton: Princeton University Press, 1988.

---------------, Theorizing the Moving Image, New York: Cambridge University Press, 1996.

Chaffin, R. "Knowledge of Language and Language About the World: A Reaction Time Study of Necessary and Invited inferences." In Cognitive Science 1979, 3.

Chase, William G. "Visual Information Processing," in Handbook for Perception and Human Performance II, ed. by Keneth Boff, Lloyd Kaufman, and James P. Thomas, NY: John Wiley and Sons, 1986.

----------------, & H. A. Simon, "Perception in Chess" Cognitive Psychology, 1973: 4.

Chatman, Seymour Coming to terms: The Rhetoric of Narrative in Fiction and Film, Ithaca, N.Y.: Cornell University Press, 1990.

-----------------------, Story and Discourse, Cornell University Press, 1978.

Chomsky, Noam Language and Mind, New York: Harcourt Brace Jovanovich, Inc. 1972.

---------------------. Rules and Representations, NY, NY: Columbia University Press, 1980.

Cook, V.J. Chomsky's Universal Grammar: An Introduction, Oxford, UK: Basil Blackwell, 1988.

Culler, Jonathan, "Fabula and Sjuzhet in the Analysis of Narrative: Some American Discussions," Poetics Today 1 (1980): 27-37.

Currie, Gregory Image and Mind: Film, Philosophy, and Cognitive Science, New York, NY: Cambridge University Press, 1995.

------------------, The Nature of Fiction, New York, NY: Cambridge University Press, 1990.

Deleuze, Gilles, Cinema 1: The Movement -Image, trans. by Hugh Tomlinson and Barbara Habberjam, Minneapolis: University of Minnesota Press, 1986.

Derrida Jacques "But, Beyond . . ." Critical inquiry, 13, (1986).

Dretske, Fred "Meaningful perception," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995.

-----------------, Seeing and Knowing, Chicago: University of Chicago Press, 1969.

Dillon, George L. Language Processing and the Reading of Literature: Toward a Model of Comprehension Bloomington: Indiana University Press, 1978.

Douglas, Ian, Film and Meaning, Perth, Australia: Continum Publications, 1988.

Eco, Umberto, A Theory of Semiotics, Bloomington: Indiana University Press, 1976.

------------------, The Role of the Reader: Explorations in the Semiotics of Texts, Bloomington: Indiana University Press, 1979.

Eikhenbaum, Boris The Poetics of Cinema, trans. By Richard Taylor, Oxford: RPT Publications, 1982.

Eisenstein, Sergei Film Form: Essays in Film theory, trans. By Jay Leyda, New York: Harcourt, Brace & World, 1949.

Elio, Renée & Francis Jeff Pelletier, "Belief Change as Propositional Update," Cognitive Science, Vol. 21 (4) 1997, pp. 419-460.

Farah, Martha J. "Dissociable Systems for Visual Recognition: A Cognitive Neuropsychology Approach," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995.

Fish, Stanley, Is There a Text in This Class? Cambridge, Mass.: Harvard University Press, 1980.
---------------"Literature in the Reader: Affective Stylistics," in (ed.) Jane Tompkins, Reader Response Criticism: From Formalism to Post Structuralism, Baltimore: Johns Hopkins University Press, 1980.

Fiske, John, and John Hartley, Reading Television, London: Methuen, 1978.

Fodor Jerry, The Modularity of Mind, Cambridge, MA: MIT Press, 1983.
---------------. "Precis of the Modularity of the Mind," in Behavioral and Brain Sciences, 8:1-5, 1985.
-------------- & Z.W. Pylyshyn, "How Direct is Visual perception? Some Reflections on Gibson's 'Ecological Approach'," Cognition 9, 1981.

French, M. Westerns: Aspects of a Movie Genre, Viking Press, 1973.

Gaudreault, André, "Narration and Monstration in the Cinema," Journal of Film and Video 39 (Spring 1987).

Gee, James Paul & Francois Grosjean, "Empirical Evidence for Narrative Structure," in Cognitive Science 8, 1984.

Genette, Gerard Narrative Discourse: An Essay in Method, trans. by Jane E. Lewin. Cornell University Press, 1972.
-------------------. Narrative Discourse Revisited, trans. By Jane E. Lewin, Ithaca, NY: Cornell University Press, 1988.

Gerrig, Richard J. Experiencing Narrative Worlds, New Haven: Yale University Press, 1993.

Gibson, J.J. "On Theories For Visual Space Perception," in G. Jansson, S.S. Bergström & W. Epstein (eds.) Perceiving Events and Objects, Hillsdale, NJ: Lawrence Erlbaum Associates, 1994.
---------------. The Ecological Approach to Visual Perception, Boston, MA: Houghton Mifflin, 1979.

Gleitman, L.R. & H. Gleitman, "A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition," Current Directions in Psychological Science, 1992:1 pp. 31-35.

Goffman, Erwin, Frame Analysis: An Essay on the Organization of Experience, Cambridge, Mass.: Harvard University press, 1974.

Gorbman, Claudia, Unheard Melodies: Narrative Film Music, Bloomington: Indiana University Press, 1987.

Greimas, Algirdas Julien, On meaning: Selected Writings in Semiotic Theory, trans by Paul G. Perron and Frank H. Collins, Minneapoils: University of Minnesota Press, 1981.

——————————, and Joseph Courtés, "The Cognitive Dimension of Narrative Discourse," New Literary History 7 (Spring 1976): 433-47.

Haber, R.N. "Control of Eye Movements in Reading" in R.A. Monty & J.R. Senders (eds.) Eye Movements and Psychological Processes, Hilsdale, N.J.: Erlbaum, 1976.

Harnad, Stevan, ed. Categorical Perception: The Groundwork of Cognition, New York: Cambridge University press, 1987.

Haviland, S.E. & Clark H.H., "What's New? Acquiring New Information as a Process in Comprehension," Journal of Verbal Learning and Verbal Behavior, 1974, 13.

Heath, Stephen, Questions of Cinema, New York: Macmillan, 1981.

Hedges, Inez, Breaking the Frame: Film Language and the Experience of Limits, Bloomington: Indiana University Press, 1991.

Heil, J. Perception and Cognition, Berkeley: University of California Press, 1983.

Hochberg J. "Toward a Speech-Plan Eye-Movement Model of Reading," in R.A. Monty & J.R. Senders (eds.) Eye Movements and Psychological Processes, Hilsdale, N.J.: Erlbaum, 1976.

——————— "Representation of Motion and Space in Video and Cinematic Displays," in Handbook of Perception and Human Performance 1, ed. by Keneth R Boff, Lloyd Kaufman, and James P. Thomas, New York: John Wiley and Sons, 1986.

——————— & V. Brooks, "The Perception of Motion Pictures," in Handbook of Perception, eds. E.C. Carterette and M. Friedman, vol.10, New York: Academic Press, 1978.

Ingarden, Roman The Cognition of the Work of Art, trans. By Crowley & Olson, Evanston: Northwestern University Press, 1973.

Iser, Wolfgang The Act of reading: A Theory of Aesthetic Response, Baltimore and London: Johns Hopkins University Press, 1978.

Jakobson, Roman, "Visual and Auditory Signs" in Selected Readings 2, The Hague: Mouton, 1971.

---------------------, "Linguistics and Poetics" in The Structuralists From Marx to Lévi Strauss, ed. by Richard and Fernande de George, Garden City, N.Y.: Doubleday, 1972.

Jauss, Hans Robert, Toward an Aesthetic of Reception, Minneapolis: University of Minnesota Press, 1982.

Johnson, Mark The Body in the Mind: The Bodily Basis of Reason and Imagination, Chicago: University of Chicago Press, 1987.

Johnson Mark H. and John Morton Biology and Cognitive Development: The Case of Face Recognition, Oxford UK & Cambridge, USA, 1991.

Johnson-Laird, P.N. Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness, Cambridge, MA: Harvard University Press, 1983.

Kalinak, Kathryn, Settling the Score: Music and the Classical Hollywood Film, Madison: University of Wisconsin Press, 1992.

Keenan J.M. & Kintsch W. "The Identification of Implicitly and Explicitly Presented Information" in W. Kintsch (ed.) The Representation of meaning in Memory, Hilside, NJ: Erlbaum, 1974.

Kincaid, James R. "Coherent Readers, Incoherent Texts," Critical Inquiry 3 (Summer 1977).

Kosslyn, Stephen M. Image and Mind Cambridge, MA: MIT Press, 1980.

Kozloff, Sarah Invisible Storytellers: Voice-Over Narration in American Fiction Film, University of California Press, 1988.

Lakoff, George "Cognitive Semantics," in Meaning and Mental Representations, eds. Umberto Eco, Marco Santambrogi and Patrizia Violi, Bloomington and Indianapolis: Indiana University Press, 1988.

--------------------, Women, Fire and Dangerous Things: What Categories Reveal About the Mind, Chicago, IL: University of Chicago Press, 1987.

de-Lauretis, Theresa and Stephen Heath, eds. The Cinematic Apparatus, New York: St. Martin's Press, 1980.

Levine, M. "You-are-here maps: Psychological considerations," Environment and Behaviour, 1983:16.

Lieberman, Philip "The Origins and Evolution of Language" in Tim Inglod (ed.) Companion Encyclopedia of Anthropology: Humanity, Culture, and Social Life, New York and London: Routledge, 1994.

MacCabe, Colin "Theory and Film: Principles of Realism and Pleasure" in Mast, Cohen & Braudy (eds.) Film Theory and Criticism, Oxford University Press, 1992.
------------------, Tracking the Signifier: Theoretical Essays: Film, Linguistics, Literature, Minneapolis: University of Minnesota Press, 1985.

Marr, David, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, San Fransisco, Calif.: W.H. Freeman, 1982.

Mayne, Judith, Cinema and Spectatorship, New York: Routledge, 1993.

McClelland, J.L., P. Smolensky, D.E. Rumelhart & G.E. Hinton "Schemata and Sequential Thought Processes in PDP Models." in Parallel Distributed Processing, Vol.1, Cambridge, MA: MIT Press, 1986.

Messaris, Paul, Visual Literacy: Image, Mind, and Reality, Boulder: Westview Press, 1994.

Metz, Christian, Film Language: A Semiotics of Cinema, trans. by Michael Taylor, Oxford University Press, 1974.
------------------. The Imaginary Signifier, trans. Celia Britton et al. Bloomington, IN: Indiana University Press, 1981.
------------------. The Semiotic Challenge, Hill & Wang, New York, 1988.

Minsky, Marvin "A Framework for Representing Knowledge" in P.H. Winston (ed.) The Psychology of Computer Vision, NY, McGraw Hill, 1975.

Nakayama, Ken & Zijiang J. He, and Shinsuke Shimojo, "Visual Surface Representation: A Critical Link Between Lower Level and Higher Level Vision," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995.

Neisser, U. Cognitive Psychology, New York: Appleton- Century-Crofts, 1967.

Neuport, Richard The End: Narration and Closure in the Cinema, Detroit: Wayne State University Press, 1995.

Nichols, Bill Ideology and the Image, Bloomington: Indiana University press, 1981.

O'Brien, E.J. M.L., Rizzella, J.E. Albrecht & J.G. Halleran, "Updating a Situation Model: A Memory Based View," Journal of Experimental Psychology: Learning, Memory and Cognition, 1998, Vol. 24, 5.

Osherson, Daniel N. (general ed). An Invitation to Cognitive Science, vol. 2 &3, Cambridge, Mass.: MIT Press, 2nd Edition, 1995.

Panofsky, Erwin "Style and Medium in the Motion Pictures," Critique 1, no, 3, January- February, 1947.

Pashler, Harold "Attention and Visual Perception: Analyzing Divided Attention," in An Invitation to Cognitive Science: Visual Cognition, Vol. 2, eds. Stephen M. Kosslyn and Daniel N. Osherson, Cambridge, MA: MIT Press, 1995.

Peirce, Charles "Questions Concerning Certain Faculties Claimed for Man," in Collected Papers, ed. Charles Hartshorne and Paul Weiss, Cambridge, MA: Harvard University Press, 1931.

Piaget, J. & B. Inhelder, The Child's Conception of Space, New York, NY: Norton, 1967.

Pinker, Steven, The Language Instinct: How The Mind Creates Language, New York: William Morrow, 1994.

Ponech, Trevor "Visual Perception and Motion Picture Spectatorship," Cinema Journal 37:1, 1997.

Posner, M.I. & S.W. Keele, "On The Genesis of Abstract Ideas," Journal of Experimental Psychology, 1968, 77, 353-363.

Pratt, Mary Louise Toward a Speech Act Theory of Literary Discourse, Bloomington: Indiana University Press, 1977.

Prince, Gerald Dictionary of Narratology, University of Nebraska Press, 1987,
-----------------, Narratology: The Form and Functioning of Narrative, New York: Mouton, 1982.

Prince, Stephen, "The Discourse of Pictures: Iconicity and Film Studies," Film Quarterly 47 (Fall 1993): 16-28.

Raffman, Diana, Language, Music and Mind, Cambridge, Mass.: MIT Press, 1993.

Rayner, K. "Foveal and Parafoveal Cues in reading" in J. Requin (ed.) Attention and Performance VII, Hilside, NJ.: Erlabaum, 1978.

Riffaterre, Michael Semiotics of Poetry, Bloomington: Indiana University Press, 1978.

Rimmon-Kenan, Shlomith Narrative Fiction: Contemporary Poetics, Routledge, 1983.

Rock, Irwin, The Logic of Perception, Cambridge, Mass.: MIT Press, 1983.

Rosch, E. "On the internal structure of perceptual and semantic categories," In T.E. Moore (ed.) Cognitive Development and the Acquisition of Language, New York, NY: Academic Press, 1973.
-----------, C. Mervis, W. Gray, D. Johnson, & P. Boyes- Braem, "Basic Objects in Natural Categories," Cognitive Psychology, 1976:8.
-----------, & C.B. Mervis, "Family resemblances: Studies in the internal structure of categories," Cognitive Psychology, 1975, 7, 573-650.

Rosen, Philip (ed.), Narrative, Apparatus, Ideology: A Film Theory Reader, New York: Columbia University Press, 1986.

Rush, Jeffrey S. "Lyric Oneness: The Free Syntactical Indirect and the Boundary Between Narrative and narration," Wide Angle 8 (1986): 27-33.

Ruthrof, Horst, The Reader's Construction of Narrative, London: Routledge & Kegan Paul, 1981.

Ryan, Marie-Laure, "Fiction as a Logical, Ontological and Illocutionary Issue," Style 1984, 18:2, 121-139.
----------------------, Possible Worlds, Artificial Intelligence and Narrative Theory, Indiana University Press, 1991.

de Saussure, Ferdinand A Course in General Linguistics, trans. Wade Baskin, New York, NY: McGraw Hill, 1966.
------------------------. "The Object of Study," in David Lodge, editor, Modern Criticism and Theory, London and New York: Longman, 1988.
------------------------. "Nature of the Linguistic Sign," in H. Adams and L. Searle (eds.) Critical theory Since 1965, Tallahassee: Florida State University Press, 1986.

Schank, R.C. & R.P. Abelson, Scripts, Plans, Goals and Understanding, Hillside, NJ: Earlbaum, 1977.

Scholes Robert "Deconstruction and Communication "Critical inquiry, 14 (1988).

------------------. "Narration and Narrativity in Film and Fiction," in Semiotics and Interpretation, Yale University press, 1982.

Silberman, Mark "Review of Holub's Reception Theory," New German Critique 33 (Fall 1984).

Silverman, Kaja, The Subject of Semiotics, New York: Oxford University Press, 1983.

Solso, Robert L. Cognition and the Visual Arts, Cambridge, MA: MIT Press, 1994.

Spence, Siegfried, Ludwig Malpass & Roy S. Kuehnken (eds.) Psychological Issues in Eyewitness Identification. Mahwah, NJ: Lawrence Erlbaum Assoc., 1996.

Sperber, Dan, and Dierdre Wilson, Relevance: Communication and Cognition, Cambridge: Harvard University Press, 1986.

Staiger, Janet Interpreting Films: Studies in the Historical Reception of Cinema, Princeton: Princeton University Press, 1992.

Sternberg, Meir, Expositional Modes and Temporal Ordering in Fiction, Baltimore, MD.: Johns Hopkins University Press, 1978.
--------------------, "Telling in Time (1): Chronology and Narrative Theory," Poetics Today 11 (Winter 1990): 901-48.

Stillings Neil A., et al. Cognitive Science: An Introduction, Cambridge, Mass: MIT Press, 1987.

Street, Brian V. & Niko Besnier, "Aspects of Literacy," in Tim Inglod (ed.) Companion Encyclopedia of Anthropology: Humanity, Culture, and Social Life, New York and London: Routledge, 1994.

Toolan, Michael J., Narrative: A Critical Linguistic Introduction, New York: Routledge, 1988.

Treisman, A. "Perceptual grouping and attention in visual search for features and for objects," Journal of Experimental Psychology: Human Perception and Performance, 1982, 8.

Tversky, B. "Distortions in memory for maps," Cognitive Psychology, 1981:13.

Uspensky, Boris A Poetics of Composition, University of California Press, 1973.

Van den Broek, Paul "The Causal Inference Maker: Towards a Process Model of Inference Generation in Text Comprehension." In D.A.

Balota, G.B. Flores d'Arcais, and K. Rayner (eds.) Comprehension Processes in Reading, Hillsdale, NY: Laurence Erlbaum Associates, 1990.

deVillers, P.A. & J.G. deVillers, "Language Development" in M.H. Bronstein & M.H. Lamb (eds.) Developmental Psychology: An Advanced Textbook, 3rd ed. Hillside, NJ: Earlbaum, 1992.

Wade Nicholas J. & Michael Swanston, Visual Perception: An Introduction, London and new York: Routledge, 1991.

Weis, Elizabeth, and John Belton, eds. Film Sound: Theory and Practice, New York: Columbia University Press, 1985.

White, Hayden, The Content of the Form: Narrative Discourse and Historical Representation, Baltimore, MD.: Johns Hopkins University Press, 1987.

Willemen, Paul, "Cinematic Discourse: The Problem of Inner Speech," Screen 22 (1981): 63-93.

Wilson, George Narration in Light: Studies in Cinematic Point of View, Baltimore, MD.: Johns Hopkins University Press, 1986.
------------------. "On Film Narrative and Narrative Meaning," in Film Theory and Philosophy (eds. Richard Allen and Murray Smith), New York, NY: Oxford University Press, 1997.

Winograd, T. "Understanding Natural Language," Cognitive Psychology, 1972:3.

Wollen, Peter Signs and Meaning in the Cinema, Bloomington: Indiana University Press, 1972.

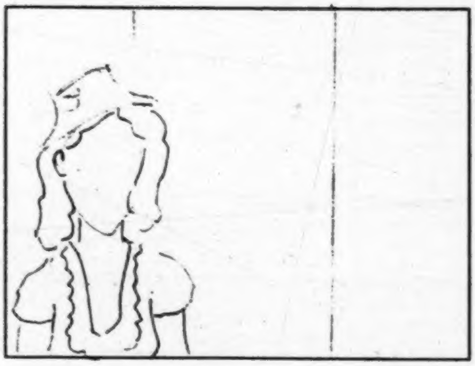
Wollheim, Richard Art and Its Objects, New York, NY: Cambridge University Press, 1980.

Worth, Sol Studying Visual Communication, Philadelphia: University of Pennsylvania Press, 1981.

Appendix #1

Rambling Rose
Rose's introduction to Dad: 1

1

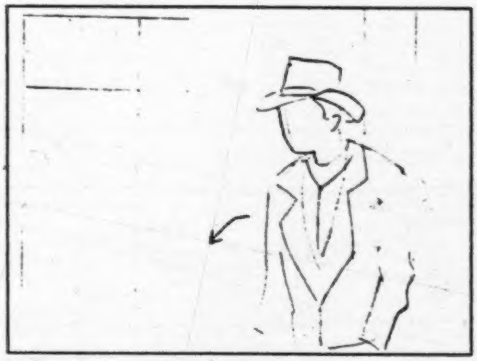


Rose seated

Mr. Hilliard (v/o): Honey you will scare the kids [. . .].

Mother laughs (v/o).

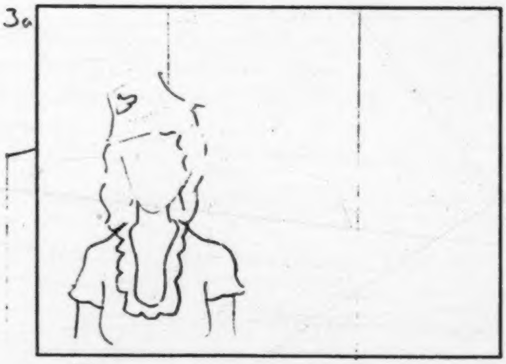
2



Camera Pans Left

Mr. Hilliard enters the frame and looks at Rose.

3a



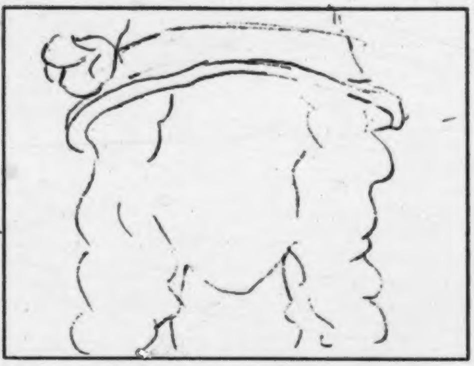
Zoom in

Rose looks off frame right, scans Mr. Hilliard up and down and gapes in pleasant surprise.

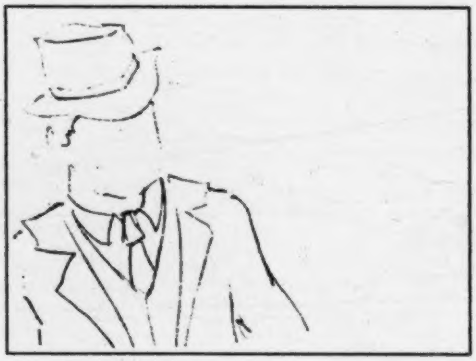
As the camera zooms in on her, romantic flute music fades in.

Rambling Rose
Rose's introduction to Dad: 1

3b



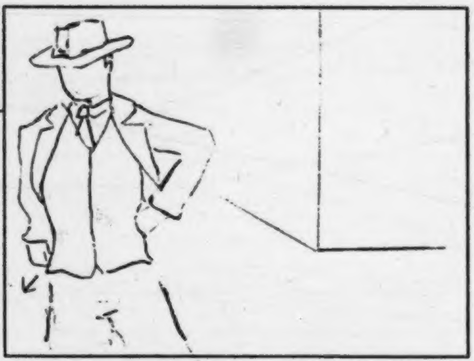
4



Mr. Hilliard: "Well, well, well. So Miss Rosebud has arrived."

Flute fades out, and the familiar film tune fades in, and will remain for the rest of the scene.

5a

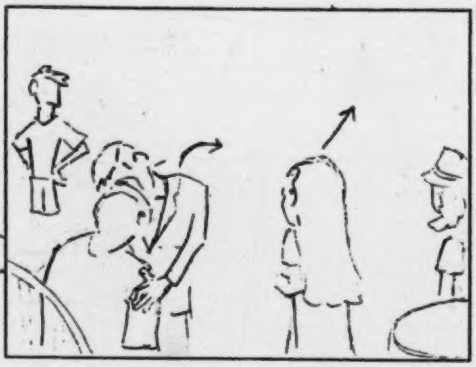


Camera is tracking Back

Mr. Hilliard moves forward, and the camera moves with him.

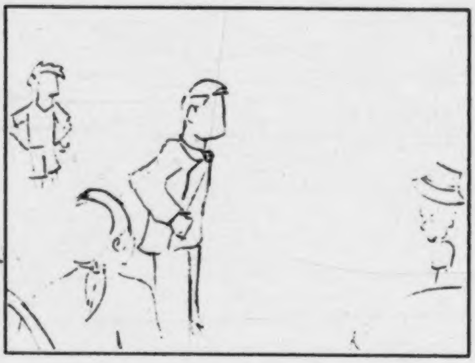
Rambling Rose
Rose's introduction to Dad: 3

5b



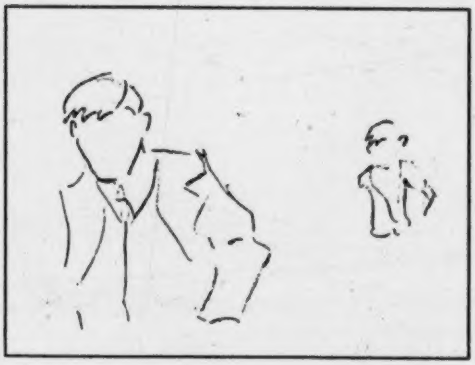
Mr. Hilliard kisses Mrs. Hilliard on the chick.

5c



Mr. Hilliard: "Rosebud."

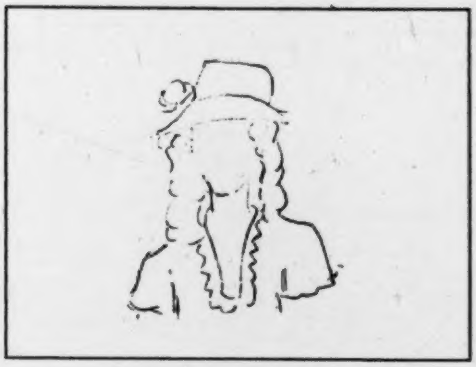
6



Mr Hiiliard (cont.): "I swear to god, you are as graceful as a capital letter S."

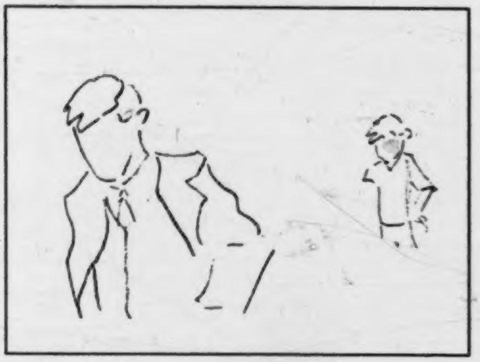
Rambling Rose
Rose's introduction to Dad: 4

7



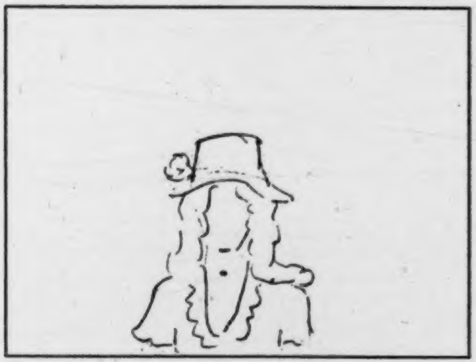
Mr. Hilliard (v/o, cont.): "You will give. . ."

8



Mr. Hilliard (cont.): ". . . a glow and a shine to these old walls."

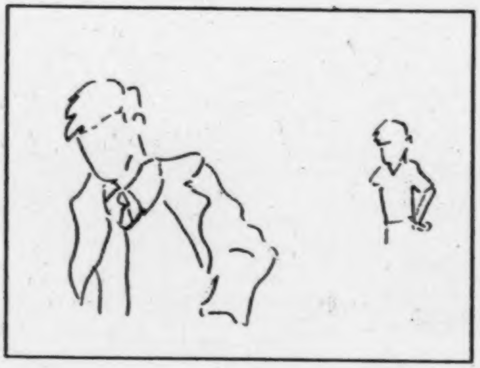
9



Rose smiles.

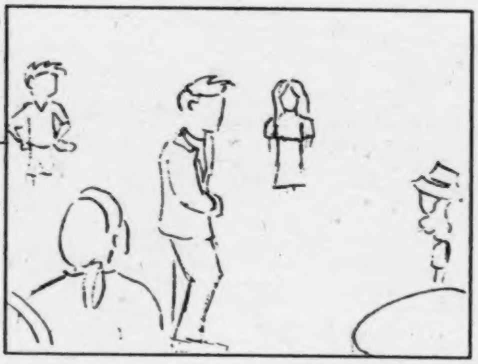
Rambling Rose
Rose's introduction to Dad: 5

10



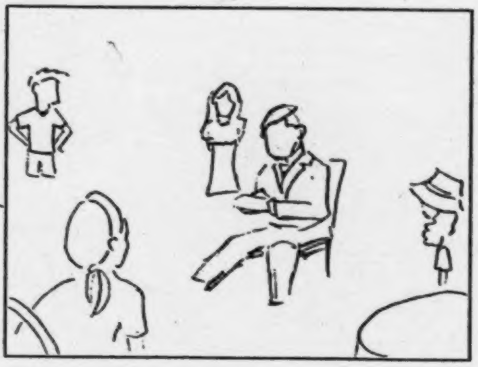
Mr. Hilliard: "Yes indeed."

11a



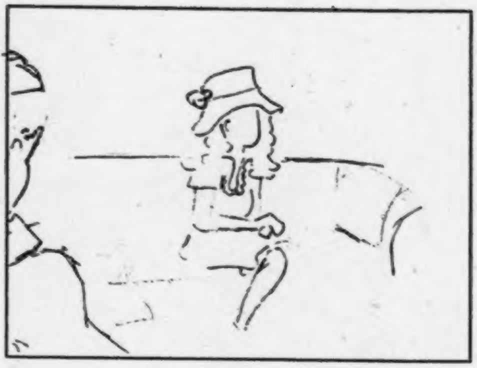
Mr. Hilliard goes to sit in his chair

11b



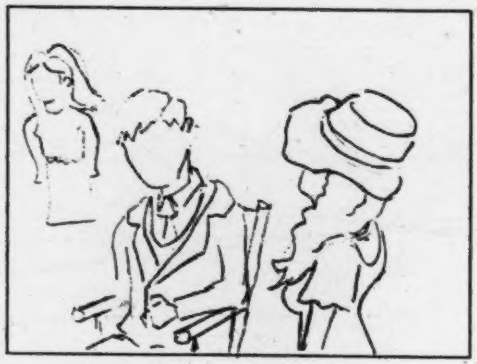
Mr. Hilliard: "Now, it is my dear wife's belief, which I accept although I do not fully understand, that to hire a person to do household work is a. . .is a criminal practice . . ."

12



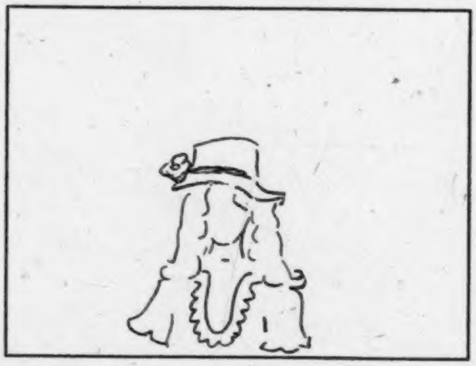
Mr. Hilliard v/o cont.): "You are therefore here as a friend, as a guest, and indeed as a member of this family."

13



Mr. Hilliard: "In love and harmony, Rosebud, in love and harmony. Do you understand me?"

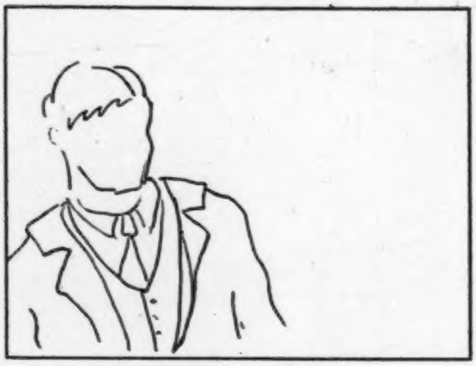
14



Rose (choked): "Yes, sir."

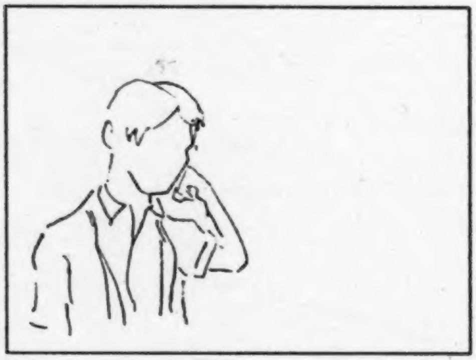
Mr. Hilliard (v/o): "I know you had some trouble in your life . . ."

15



Mr. Hilliard (cont.): "Those scoundrels in Birmingham trying to lead you astray, and so on and so forth."

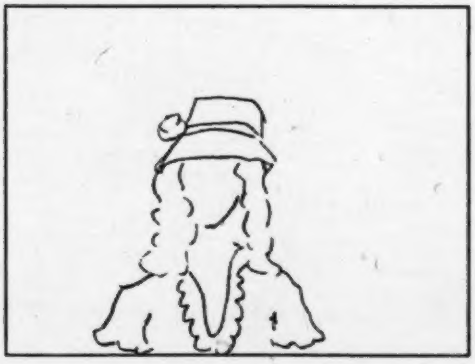
16



Buddy watches intently

Mr. Hilliard (v/o, cont.): "life can be very cruel"

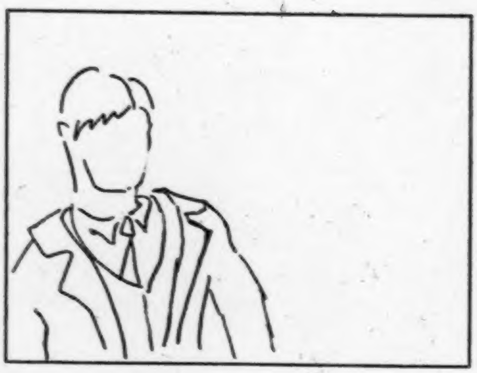
17



Mr. Hilliard (v/o, cont.): "to a young girl."

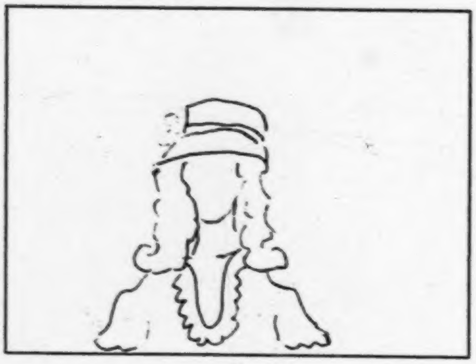
Rambling Rose
Rose's introduction to Dad: 8

18



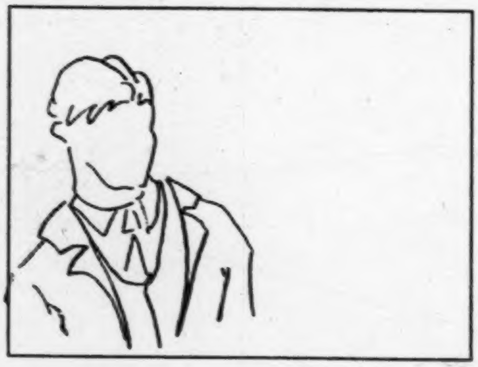
Mr. Hilliard (cont.): "I know you had a hard time, but I hope and believe that you found a safe haven in this house, honey."

19



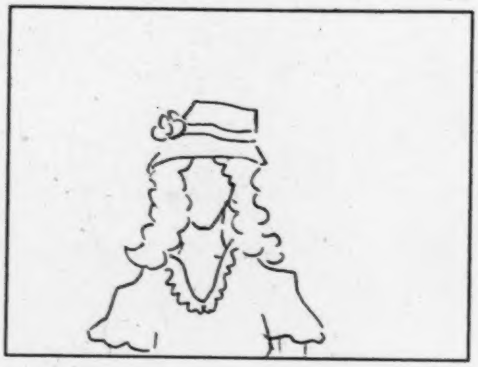
Mr. Hilliard (v/o, cont.): "Welcome to our home Rosebud."

20



Mr. Hilliard (cont.): "We welcome you from the heart and hope you are happy here."

21

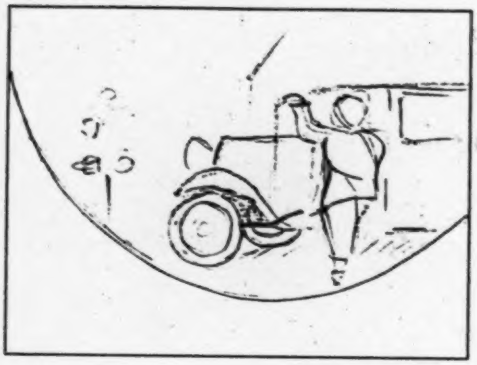


Rose (touched): "Thank you, sir."

appendix #2

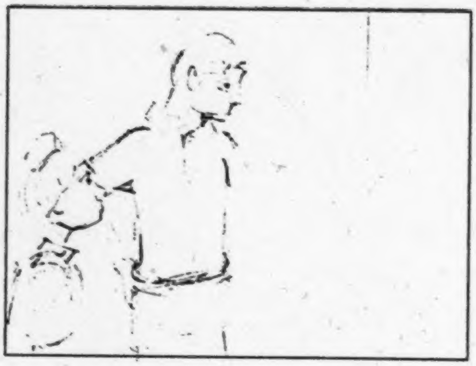
Rambling Rose
Initiation to patriarchy: 1

1



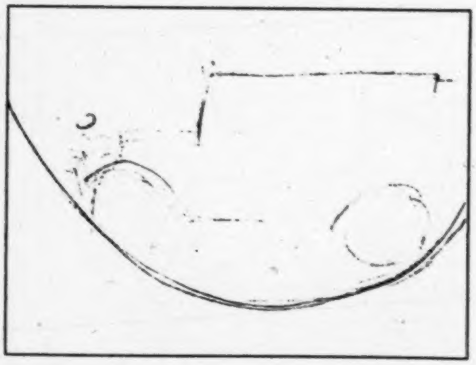
Mom is leaving to her lecture

2



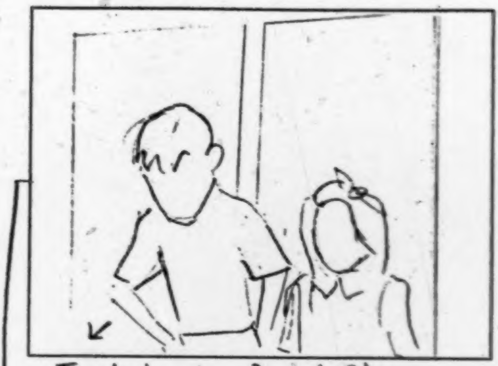
Doll and Buddy watching

3



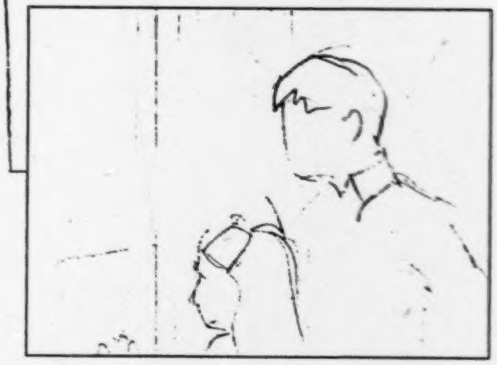
Rambling Rose
Initiation to patriarchy: 2

4



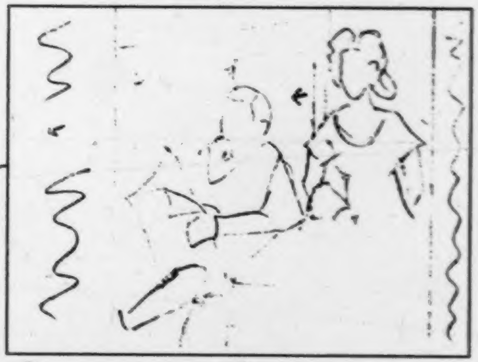
Track back Pan left

Buddy and Doll move quietly



To the cracked-open living room door

5

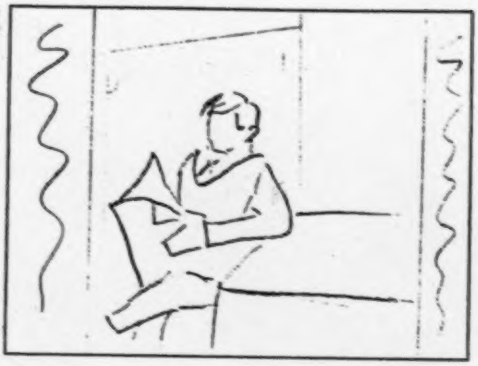


Rose leaves frame

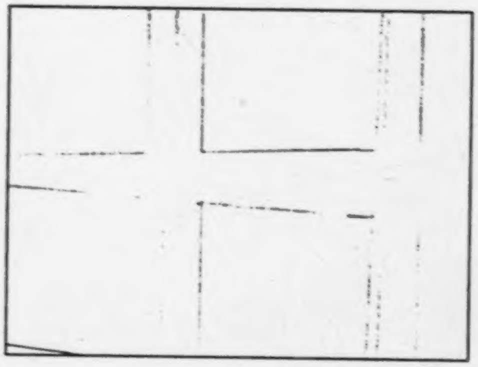
Dad is reading the paper, murmuring to himself about science revelations about the moon.

Rambling Rose
Initiation to patriarchy: 3

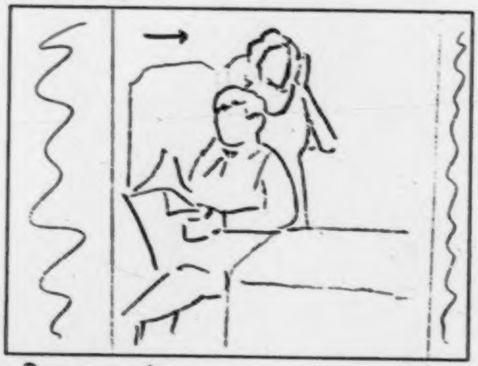
5



6

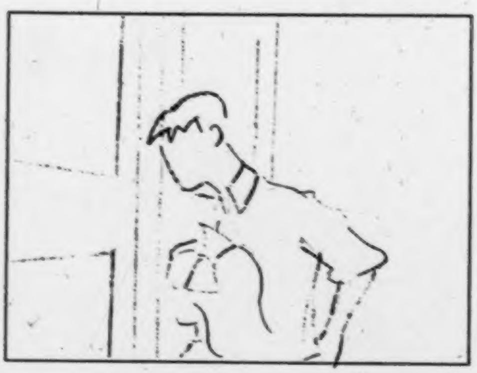


7



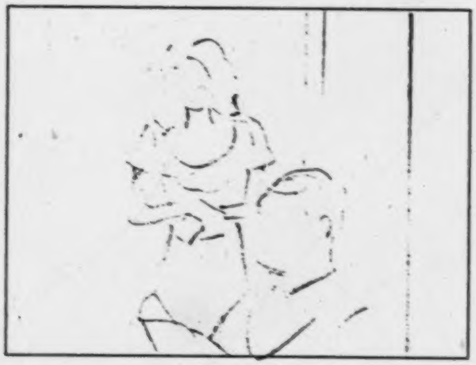
Rose returns

8

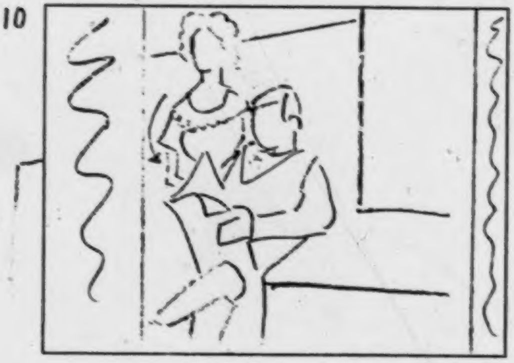


Doll: "I think she's going to kiss him."
Buddy: "Oh, at least."

9

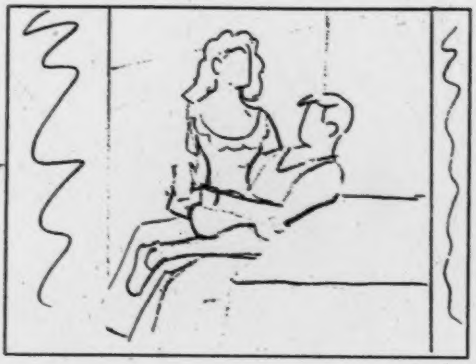


10



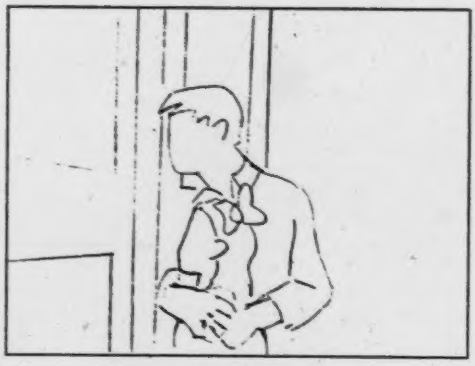
Rose jumps into Mr. Hilliard's lap, while saying: "Oh God.

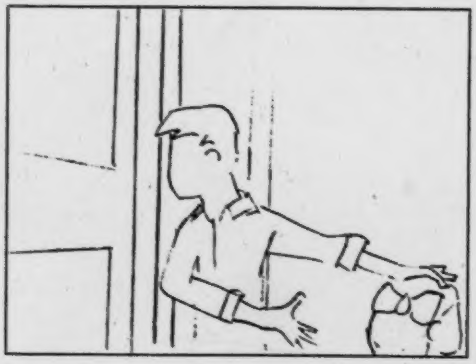
Rambling Rose
Initiation to patriarchy: 5



Rose: "I love you, I love you. I tried, but I can't help it. Please kiss me. Will you kiss me?

Dad: "Rose, Rose, get off my lap."

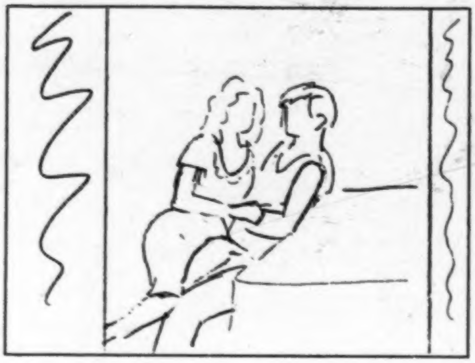




Buddy pushes Doll out of the way, blocking her view.

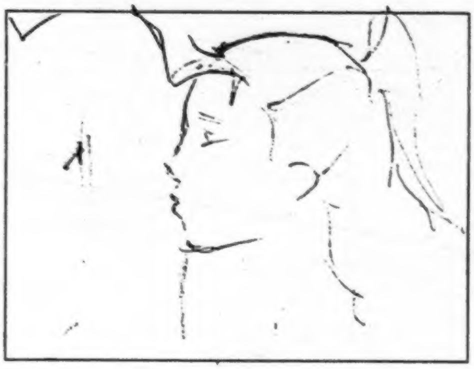
Rambling Rose
Initiation to patriarchy: 6

12



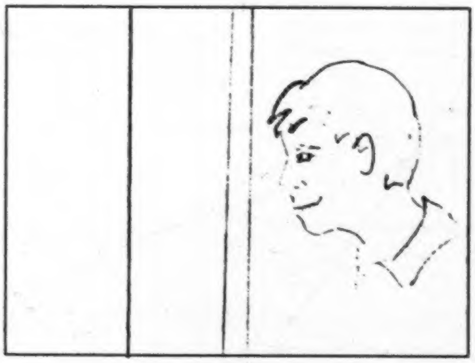
Dad: "Now calm down, calm down, the children will hear. Let's talk about it. . .let's, let's discuss it."
Rose (sobbing):"Please, please."

13



Doll: "What are they doing?"

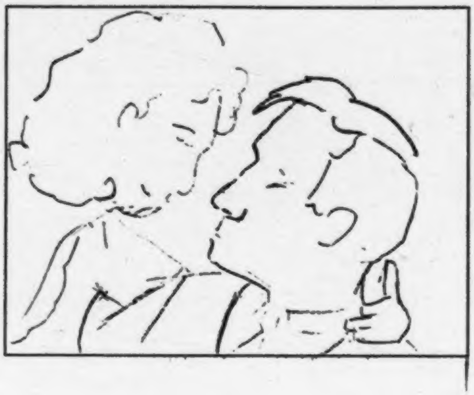
14



Buddy: "They are discussing it."

Rambling Rose
Initiation to patriarchy: 7

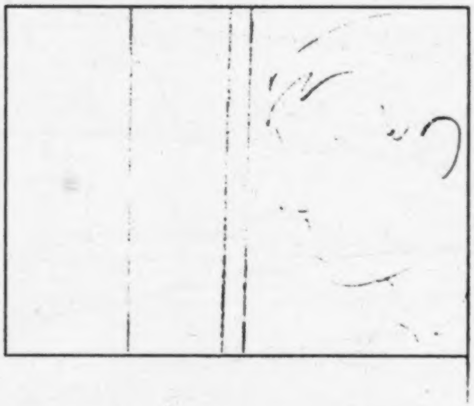
15



Dad: ". . . but I only kiss Mrs. Hilliard"
Rose: "But I love you."
Dad: "You don't."
Rose: "I do. Please just kiss me once."
Dad: "If I kiss you once you won't ask anymore?"
Rose: "Just once."
Dad: "But I only kiss Mrs. Hilliard on the moth."

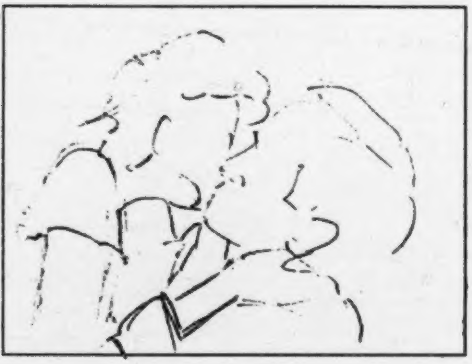
They are kissing on the mouth.

16



Buddy (quietly): "They are kissing."

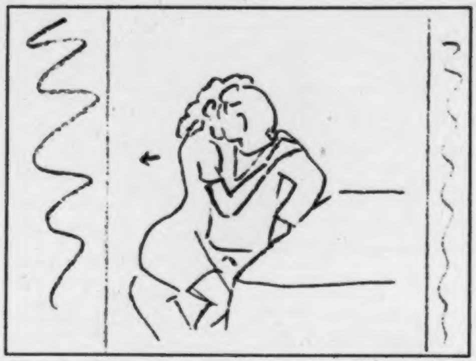
17



Rose and Dad are passionately kissing.

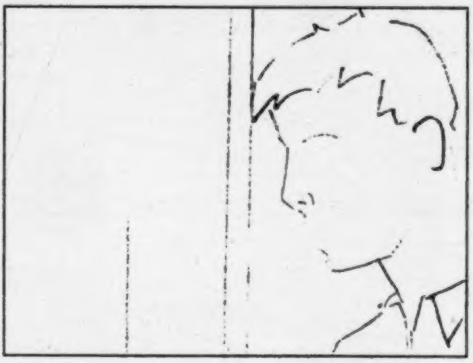
Rambling Rose
Initiation to patriarchy: 8

18

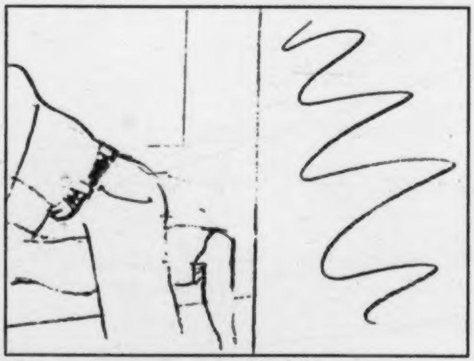


Both leave frame to the left

19

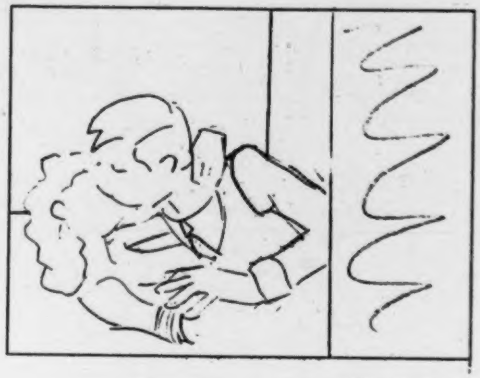


20

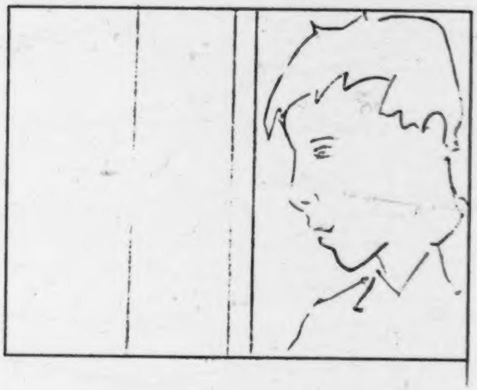


Camera pans left

Rambling Rose
Initiation to patriarchy: 9

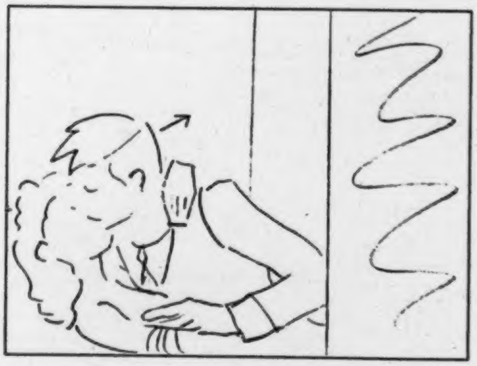


21



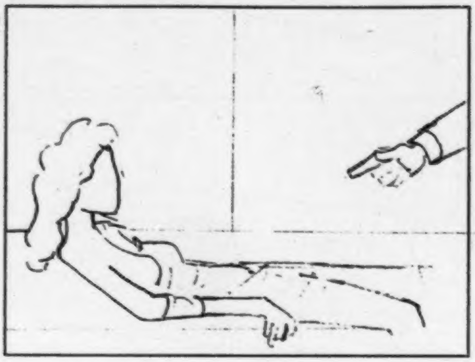
Doll (v/o): "What's happening now?
Buddy (excited): "The tity is out."

22



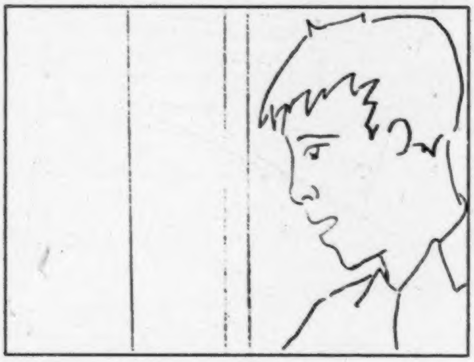
Dad gets a hold of himself and pulls away from Rose.

23

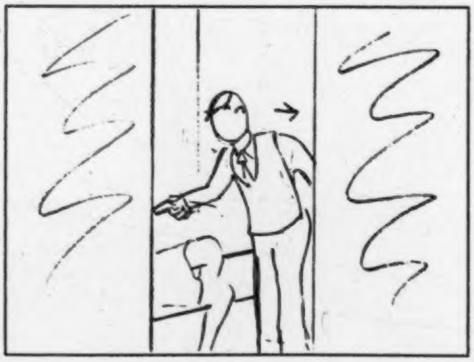


Dad: "Put that damn tit away."

24

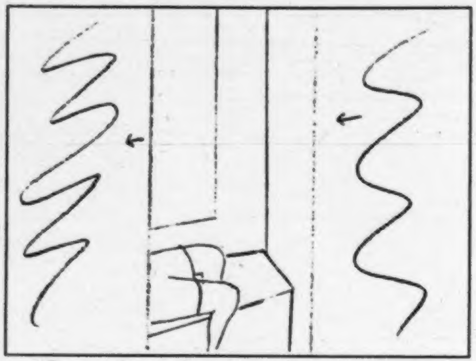


25a



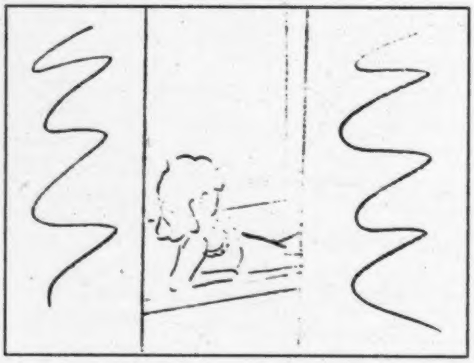
Dad: "Replace that damn tit. Damn. I am making a fool out of myself."

25b



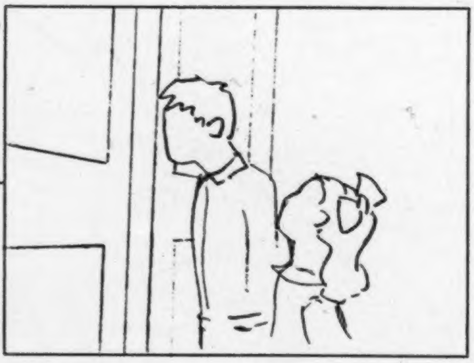
Door Pans Left

25c

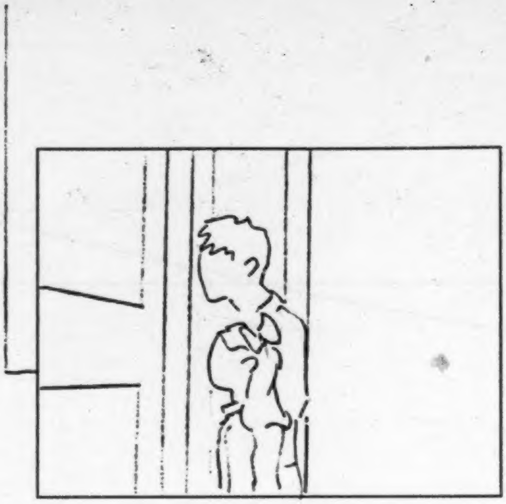


Rose is sobbing.

26

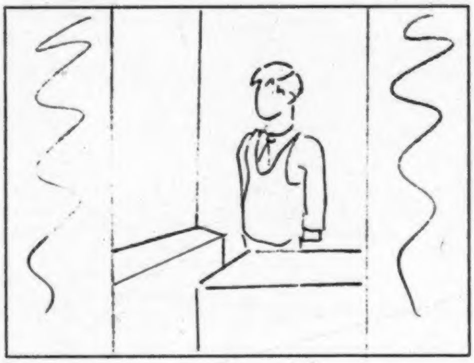


Buddy: "Oh, she put it away."



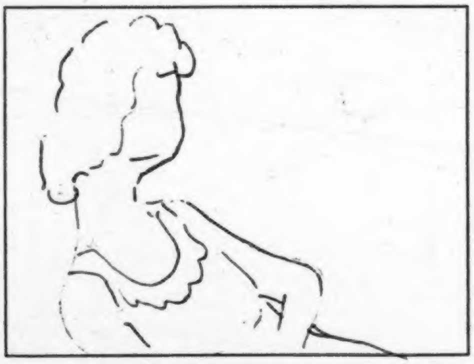
Buddy allows Doll back by the door, where she can see things for herself.

27



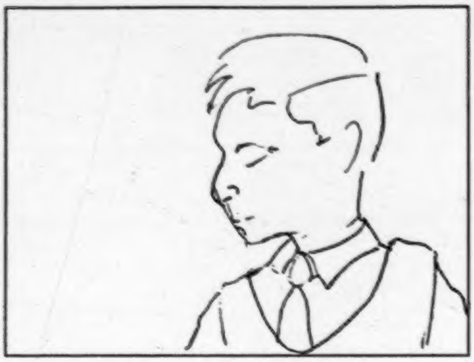
Dad: "Now a man is supposed to be a fool like this, but a woman is supposed to have some control and sense. What is the matter with you?"

28



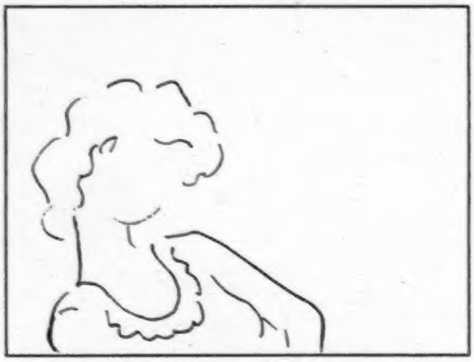
Rose: "I'm sorry Mr. Hilliard, I can't help it but I love you.

29



Dad: "Rose, Rose, Rose, you poor thing. You said yesterday that you love Mrs. Hilliard."

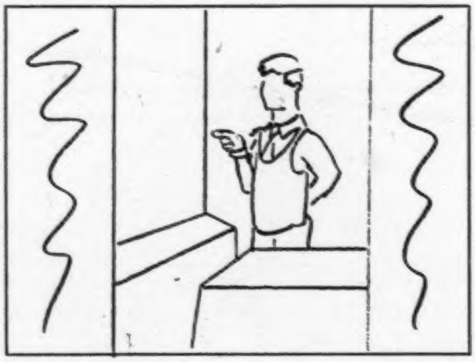
30



Dad (v/o): "Is this a way to repay her?"

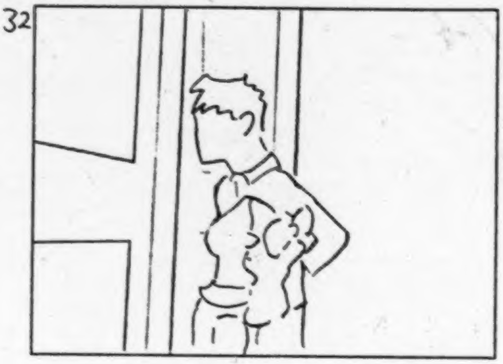
Rose is crying.

31

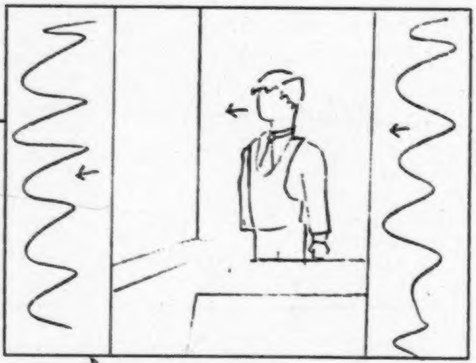


Dad (cont.): Don't you know she will fly to your defense if anyone tried to hurt you? [. . .]"

32



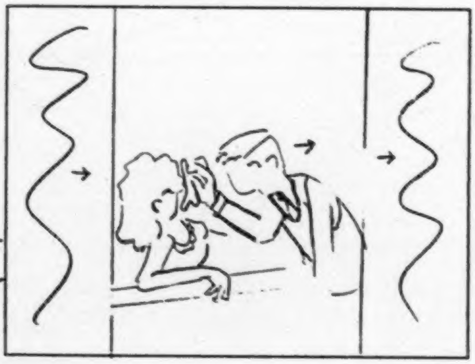
33a



Door pans left

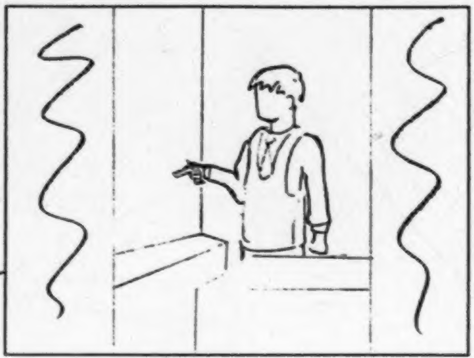
Dad: "Don't cry, honey. Don't cry"

33b



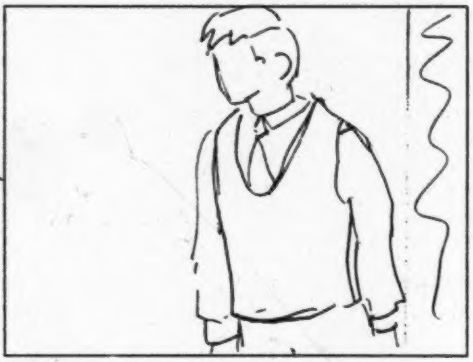
Dad hands Rose a handkerchief, and wipes her tears.

33c



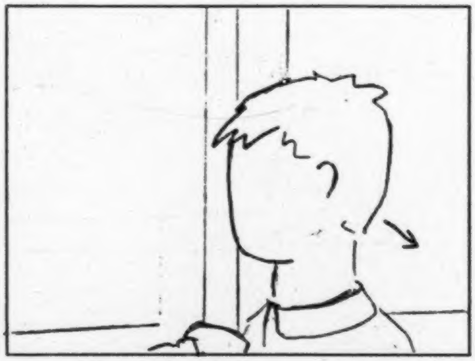
Dad: "I will worn you, do you hear me?

33d



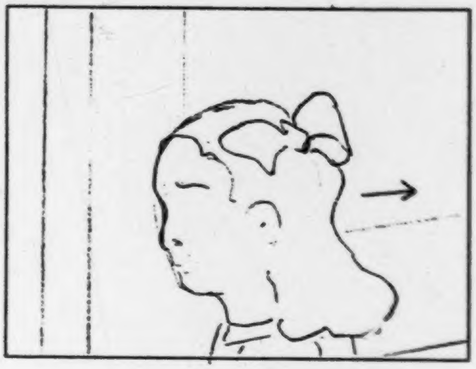
Dad (with pathos): "I am standing here at the Mapalay (sp?) and the Persians shall not pass!
Now, get your tail out of here, and go wash the dishes. Go on!!"

34

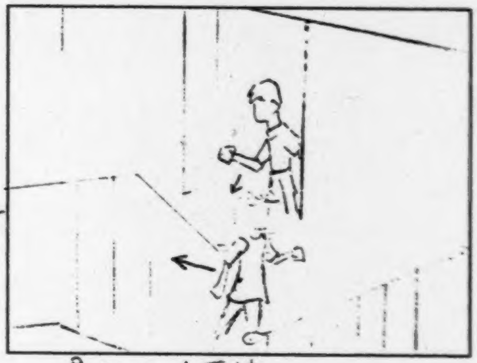


Dad hands Rose a handkerchief, and wipes her tears.

35

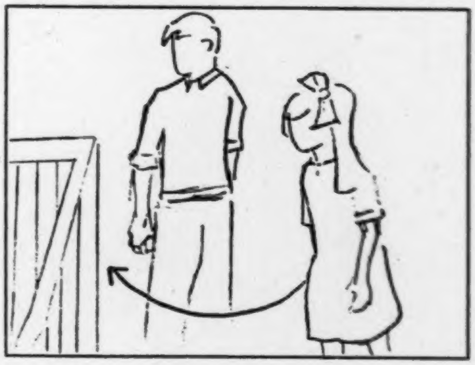


36



Kids run up the stairs before they are revealed to have been spying.

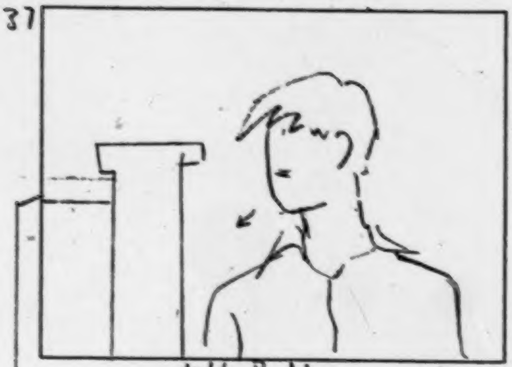
Pan and Tilt



Doll: "Weren't daddy great? I bet he wanted to kiss her some more, don't you think?"
Buddy: "He probably was just scared that mother would come home early and catch him with Rose."
Doll: "Buddy, sometimes you make me sick!"

Doll pushes past Buddy and out of the frame.

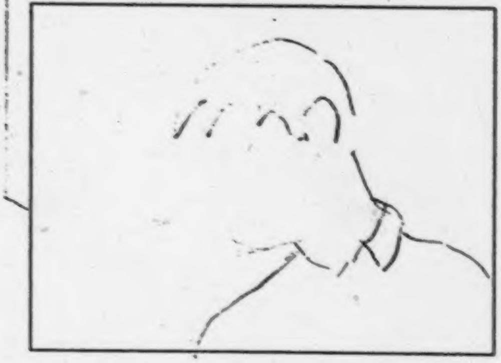
37



tilt Buddy moves in to a medium close up

Buddy (breathing excitedly):
"Oh baby Doll."

Melodic music fades in.

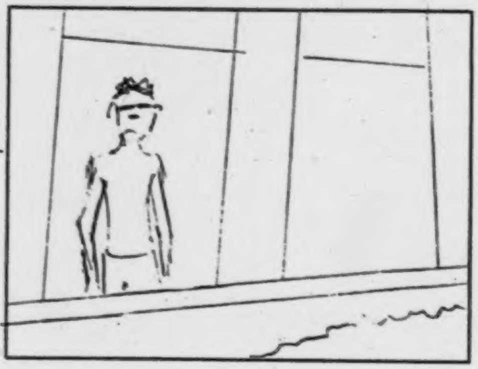


Buddy (to himself): "Mapalay. . .
The Persians shall not come . . ."

Appendix #3

Dead Poets Society
Suicide: 1

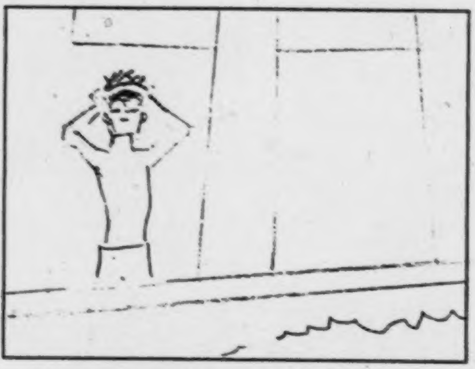
1.



Eerie music fades in

Neil picks-up the crown of thorns.

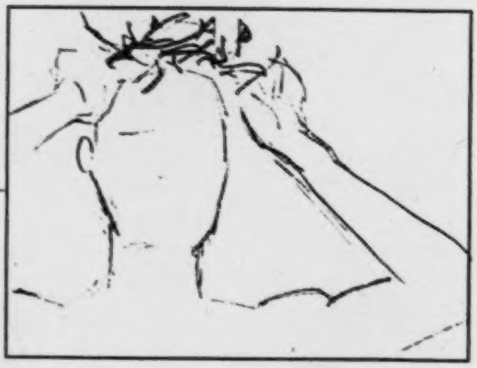
1a



camera zooms slightly

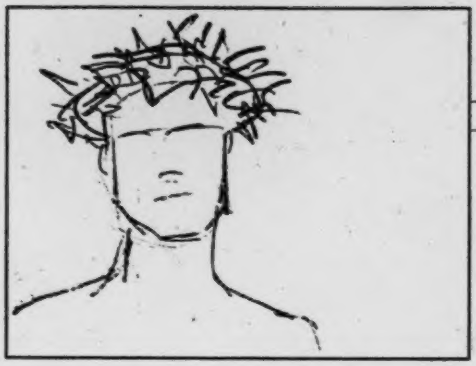
He puts it on his head

2a



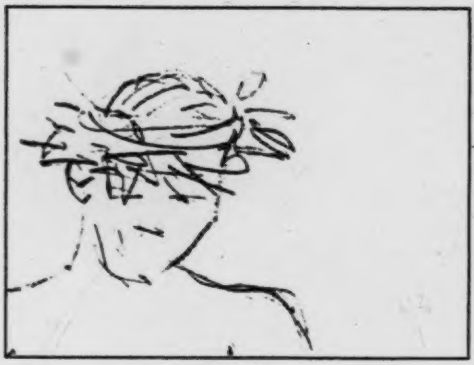
Dead Poets Society
Suicide: 2

2b



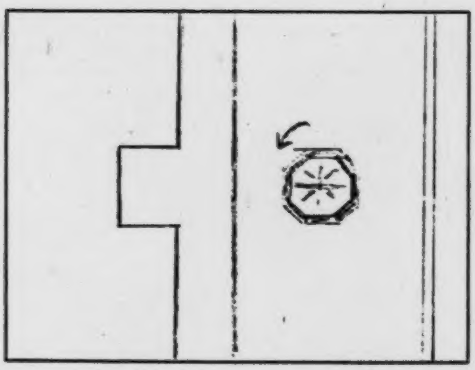
Neil closes his eyes.

2c



He looks down.

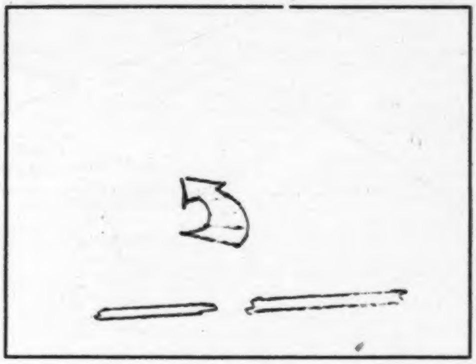
3



The door knob moves

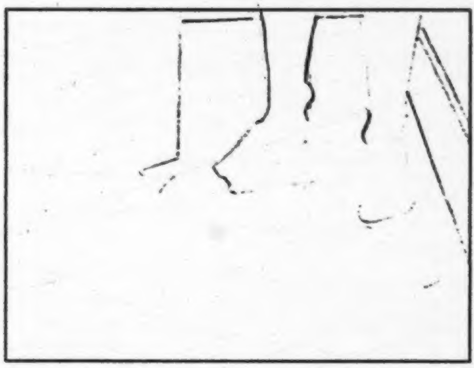
Dead Poets Society
Suicide: 3

4a



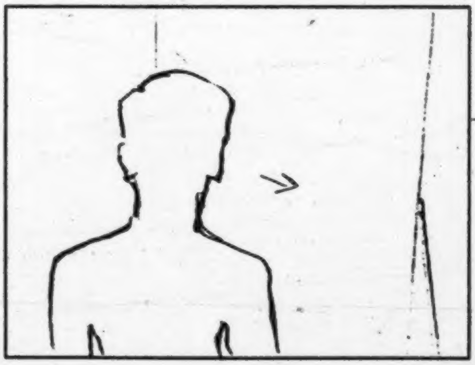
Door opens

4b



Back light shows feet

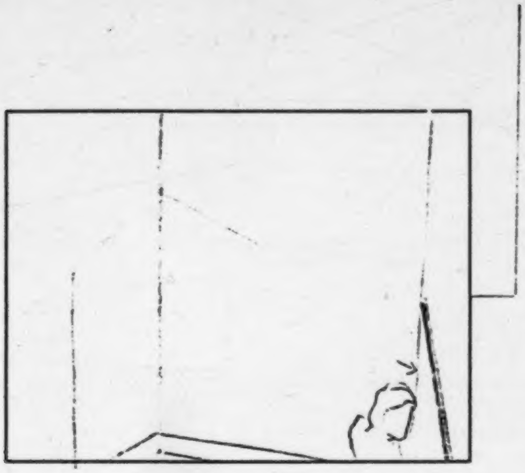
5a



Neil descends stairs.

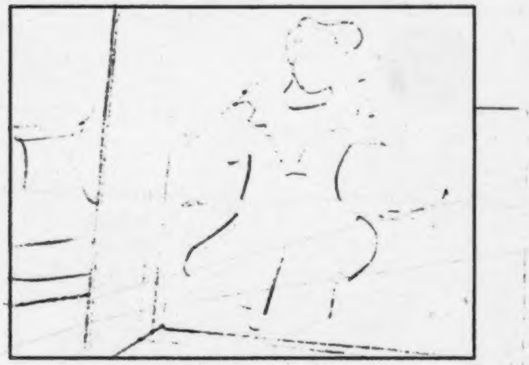
Dead Poets Society
Suicide: 4

5b



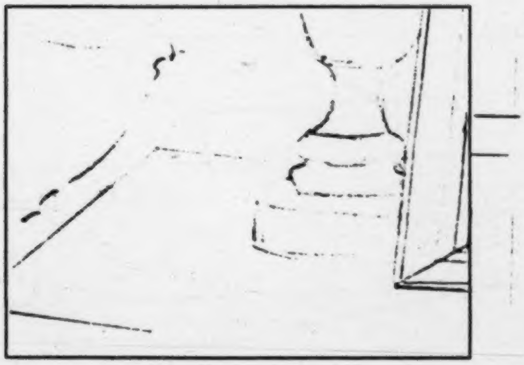
6a

Obscured and dimmly lit objects are slowly scanned through a pan movement.



Camera pans left
←

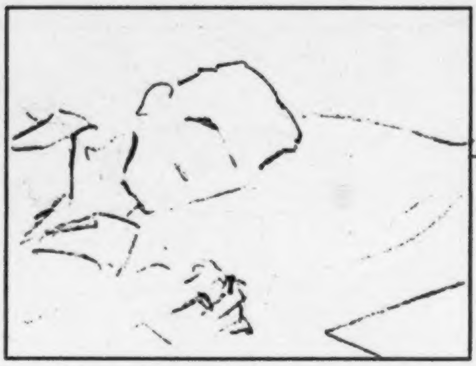
6b



Camera continues to pan

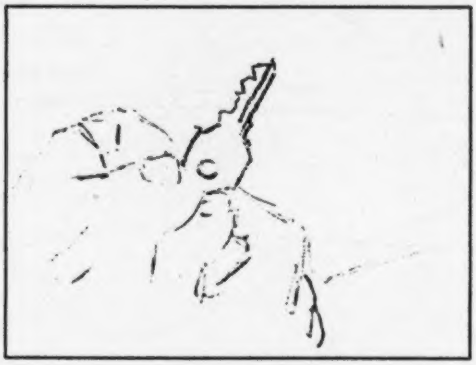
Dead Poets Society
Suicide: 5

6c

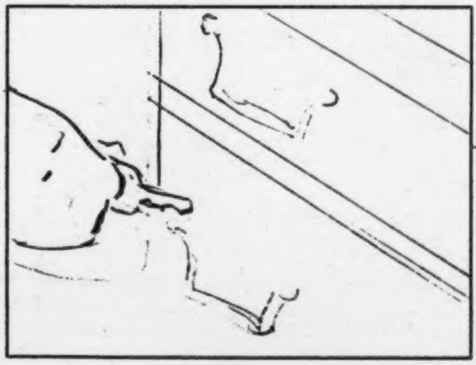


Neil's father (Tom) is asleep.

7

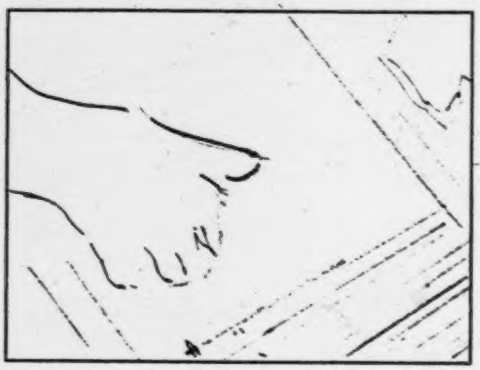


8a



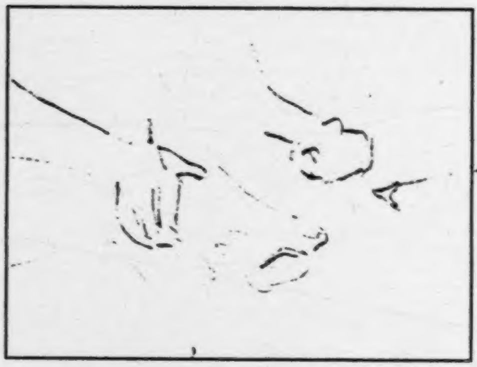
Dead Poets Society
Suicide: 6

8b



Hand reaches into drawer

9a

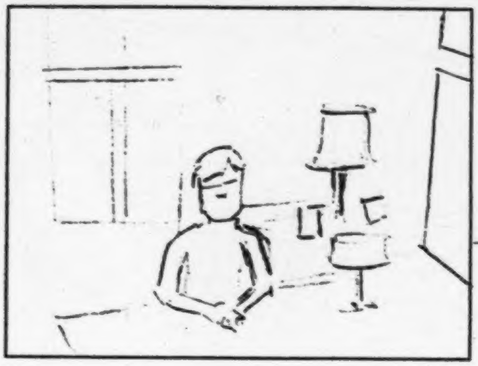


Wrapped object is pulled out.

Camera zooms out
+ tilt-up

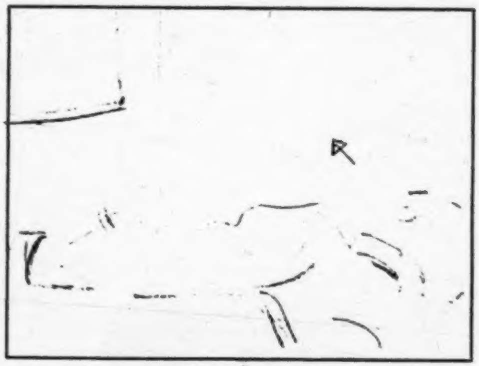
Dead Poets Society
Suicide: 7

9b



Neil is sitting at desk.

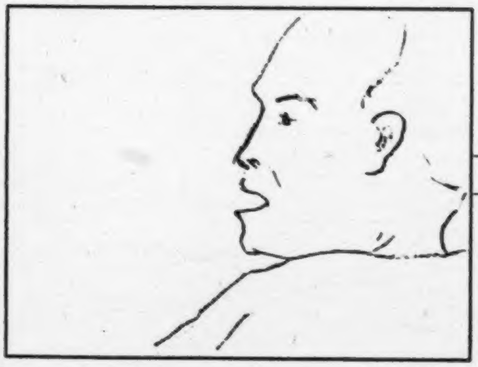
10a



Eerie music ends abruptly
while the father wakes-up
suddenly as if from a bad dream

Camera zooms in

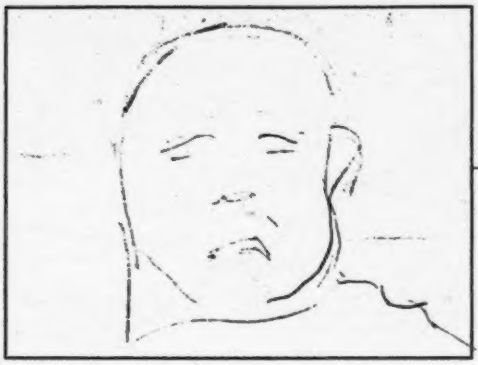
10b



The father breathes heavily
And murmurs, "this sound,"
Mother (in sleepy v/o):
"What sound?"

Dead Poets Society
Suicide: 8

10c

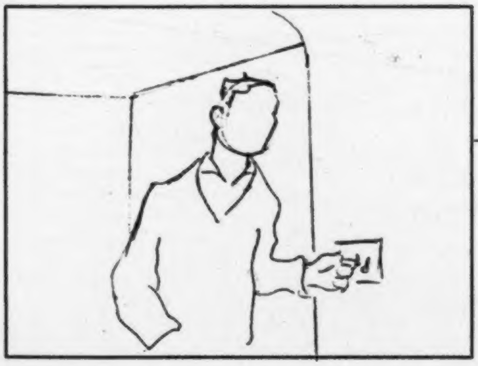


Father turns on bed-side lamp

11a

Father gets up and walks towards the camera. Mother turns on bed-side lamp.

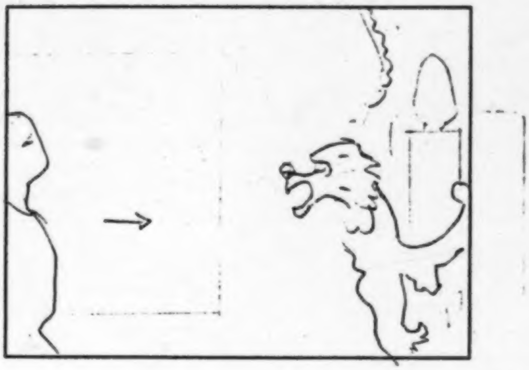
11b



Father turns on hallway light.
Mother (v/o): "What's wrong?"

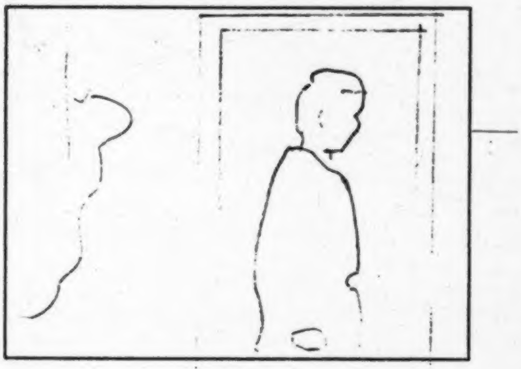
Dead Poets Society
Suicide: 9

12a



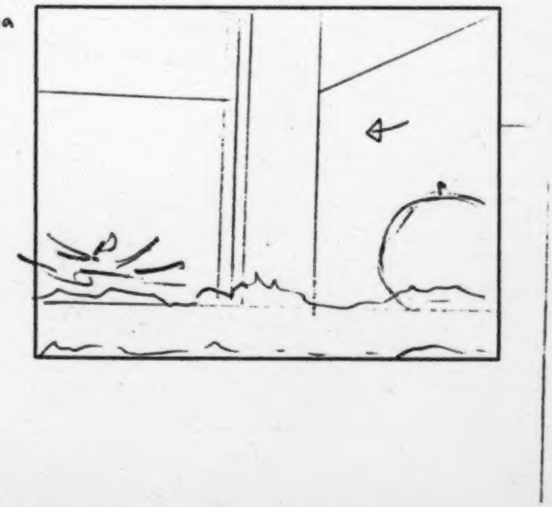
12b

Father knocks on the door and calls "Neil?"
He opens the door.



13a

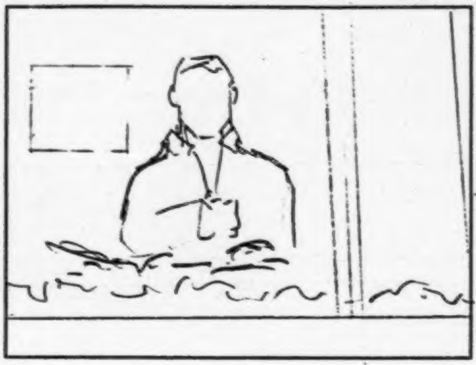
Father enters the room



Dead Poets Society
Suicide: 10

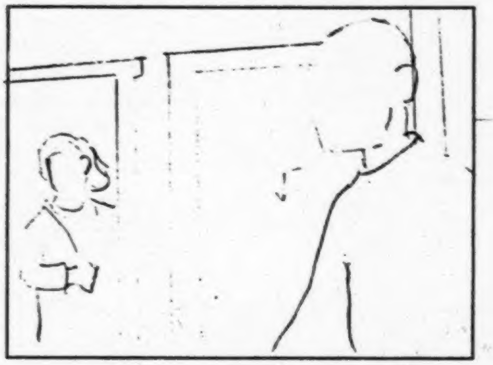
13b

Father notices the open window (and wreath?) and leaves the room.



14a

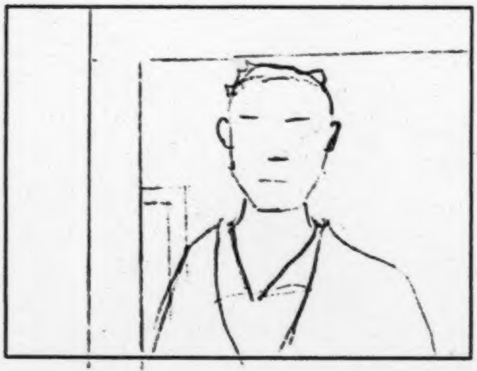
Mother (at door):
"Tom what is it?"



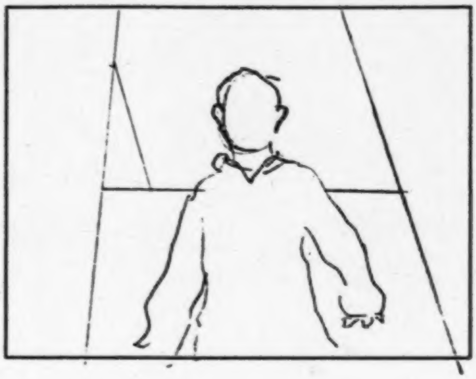
Camera tracking back

14b

Mother (v/o): "what's wrong?" Father walks down hall, camera Ahead of him at low angle. He is Silhouetted.

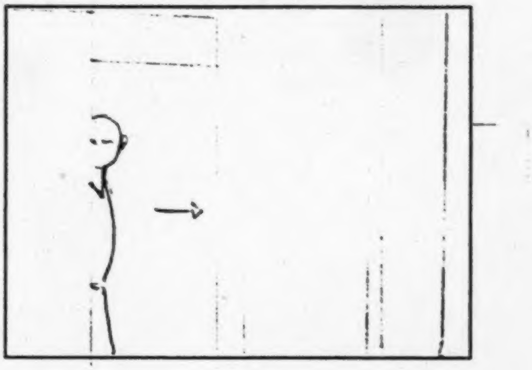


15



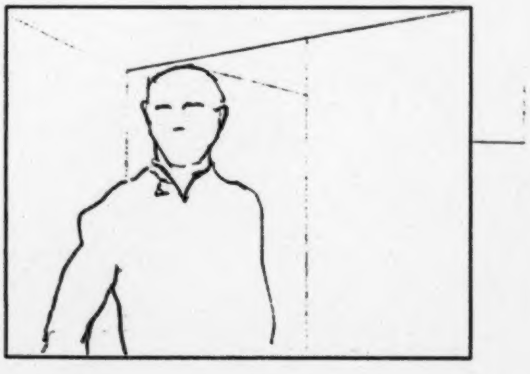
Father walks downstairs.
Mother (v/o): "Neil?!"

16a



Father: "Neil?"
Mother(v/o): "I'll look outside. Neil. . ."

16b

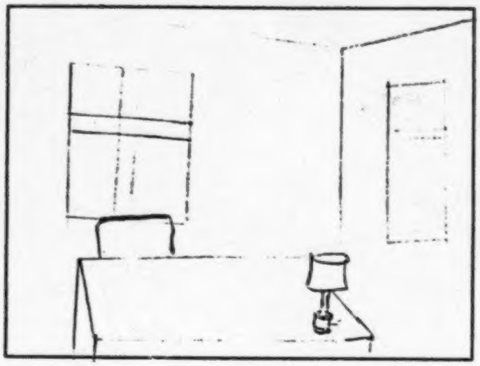


Father turns on the light.

Dead Poets Society
Suicide: 12

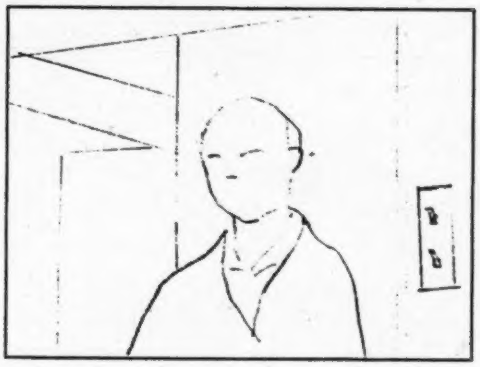
17

Wide shot of office



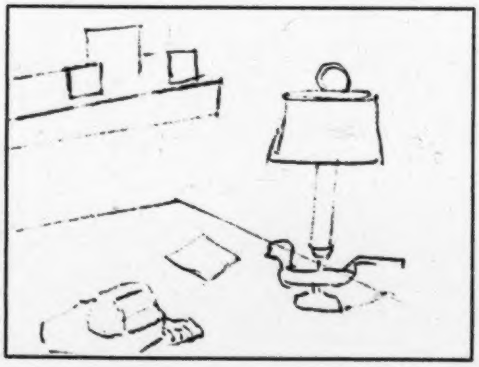
18

Father sniffs the air.



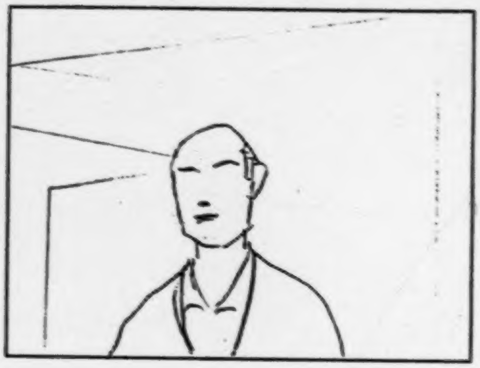
19

Medium shot on desk.



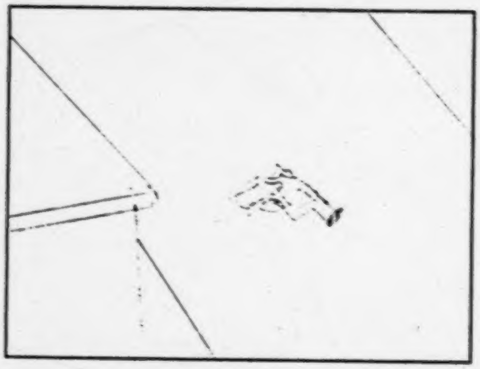
Dead Poets Society
Suicide: 13

20



Father makes a step to his right
And looks.

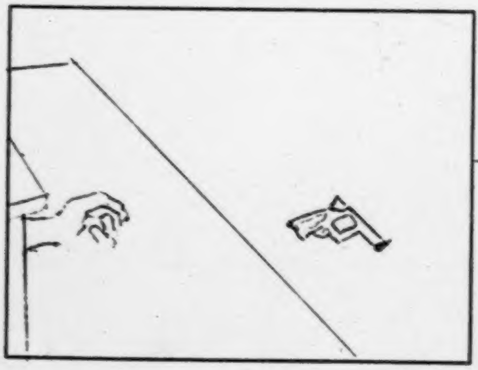
21a



Eerie music comes in.

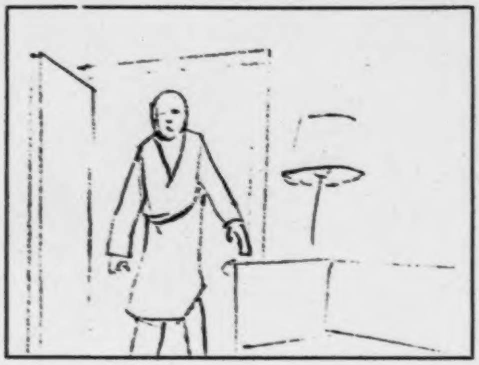
Camera pans right

21b



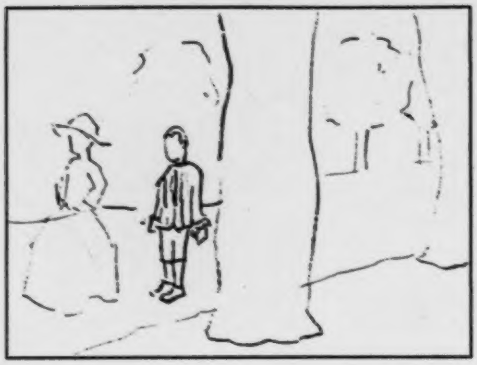
Dead Poets Society
Suicide: 14

22



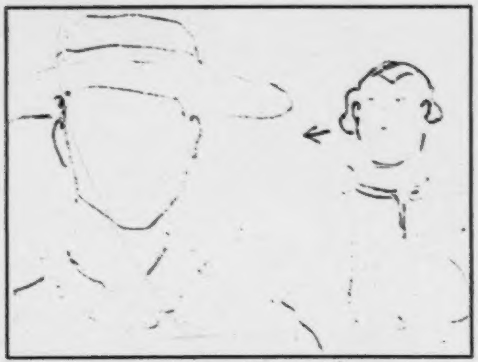
Slow motion shot.
Father moves forward.

1



Valmont is asking for a favour.

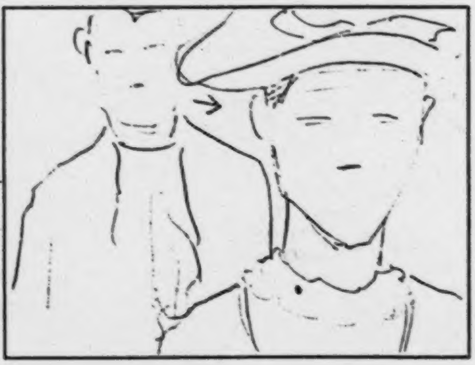
2a



Camera tracks back & pans with Valmont.

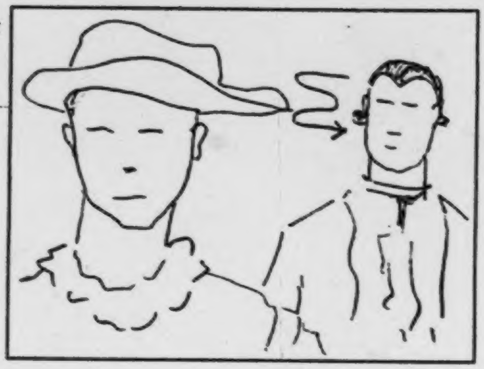
Mme. De Tourvel refuses, and they argue. Throughout the shot Valmont walks behind her changing sides from right to left of frame constantly. Mme. De Tourvel walks straight forward.

2b

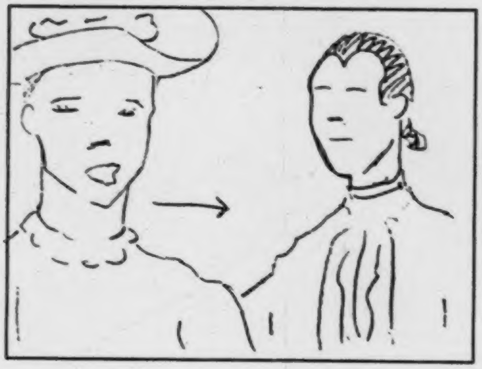


Dangerous Liasions
Seduction #1: 2

2c

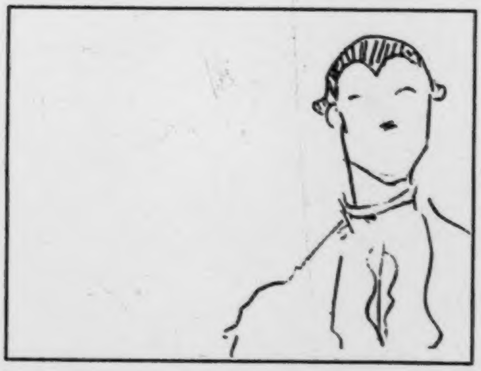


2d



Mm. De Tourvel leaves frame, cutting in front of Valmont.

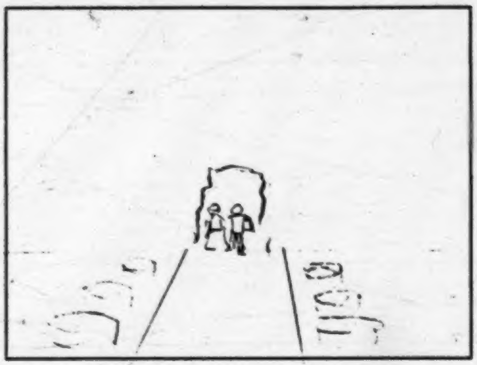
2e



Camera lingers on Valmont and the empty space on his left.

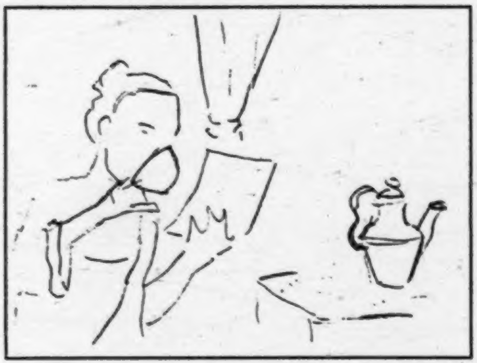
Dangerous Liasions
Seduction #2: 1

1



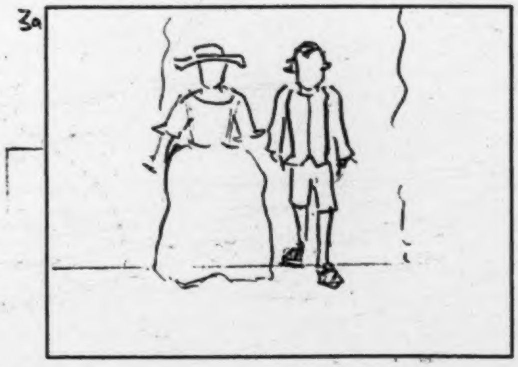
Long shot. Valmont (v/o): "We go for a walk together every day. Little further every time, down the path that has no turning."

2



Marquise De Merteuil reads Valmont's letter. Valmont (v/o cont.): "She accepted my love. I accepted her friendship."

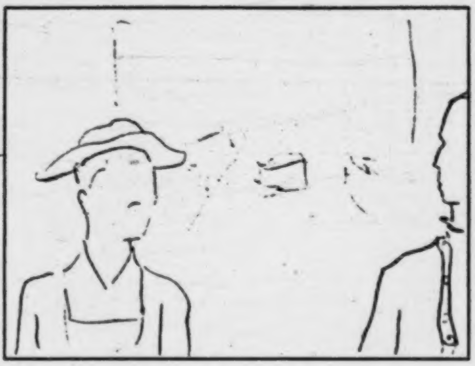
3a



Valmont (v/o cont.): "We are both aware how little there is to choose between them."

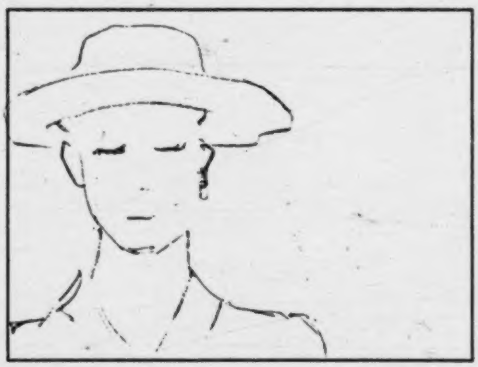
Dangerous Liasions
Seduction #2: 2

3b



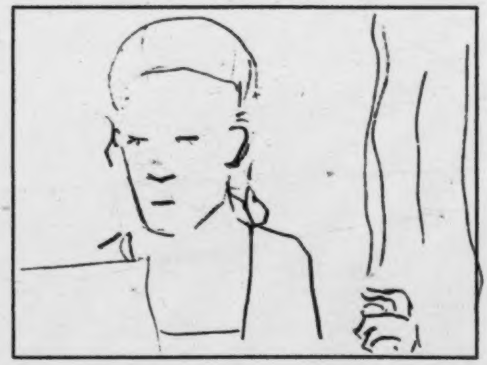
Valmont: "I wish you knew me well enough to recognize how much you've changed me. [. . .] I've become the soul of consideration, conscientious [. . .], more celibate than a monk."

3c



Mme. De Tourvel laughs: "More celibate. . ."

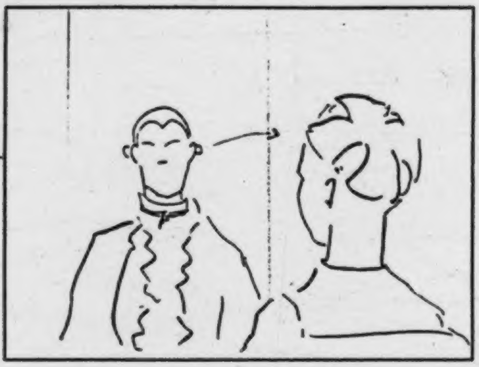
4



Marquise De Merteuil reads the letter. Valmont (v/o cont.): "I feel she is inches from surrender. Her eyes are closing.

Dangerous Liasions
Seduction #3: 1

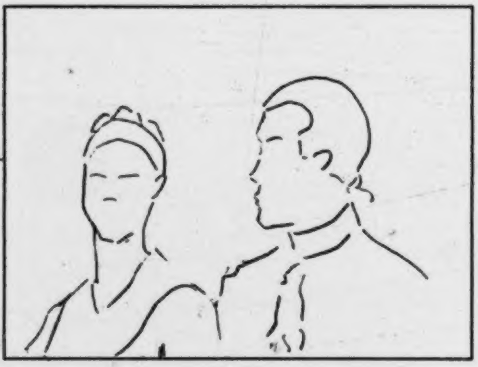
1a



camera pans right

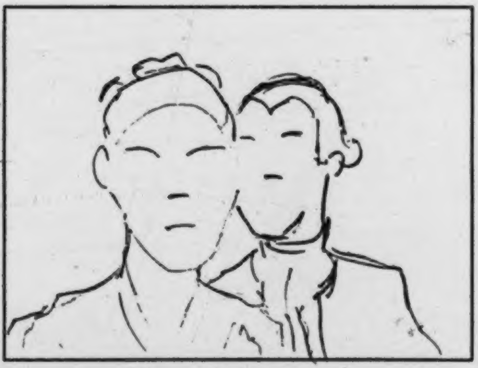
Valmont closes the door behind him.

1b



Valmont: "I've completed my business here, but I'm not sure I'll be able to bring myself to leave."

1c



camera pans right

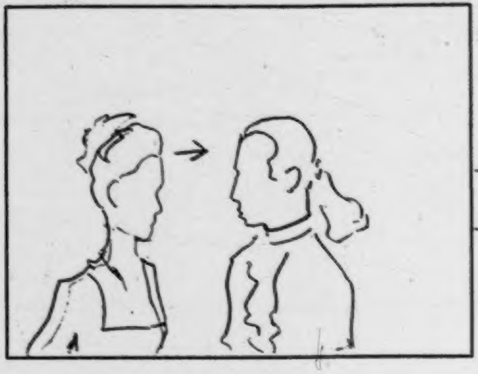
Mme. De Tourvel: "Oh please. You must."

[. . .]

Dangerous Liasions
Seduction #3: 2

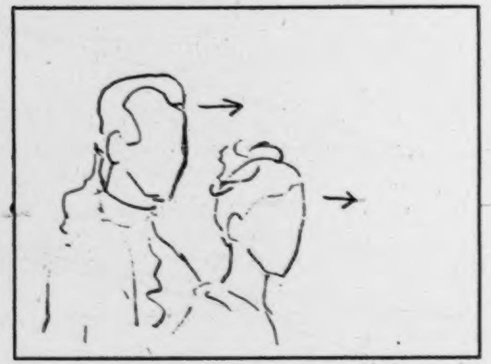
1d

Valmont: "Forgive me if I say it; I don't want your gratitude. What I want is altogether deeper."



1e

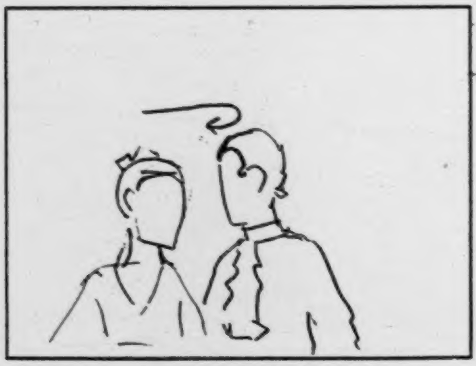
Mme. De Tourvel leaves frame to right. Valmont and camera follow. Mme. De Tourvel: "I know God is punishing me for my pride. I was so certain nothing like that could ever happen."
Valmont: "Nothing like what? You mean love?"



camera pans right

1f

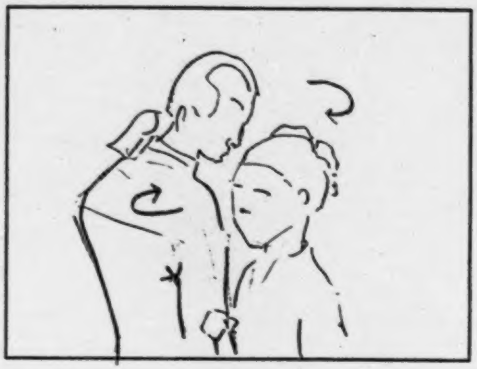
Mme. De Tourvel (distressed): "I can't."



camera pans left
and tracks in

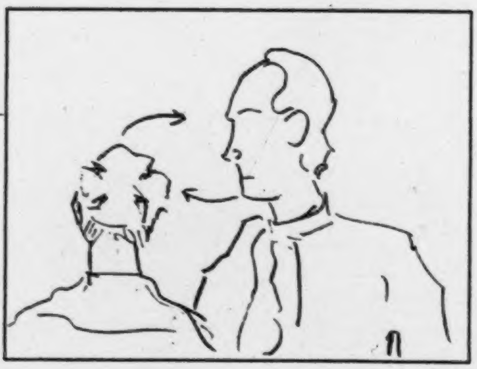
Dangerous Liasions
Seduction #3: 3

1g

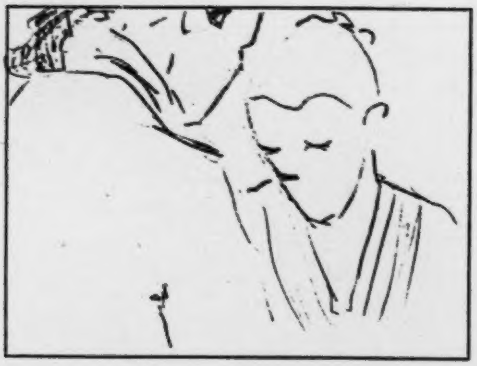


Valmont: "But I must know. I must know."

1h



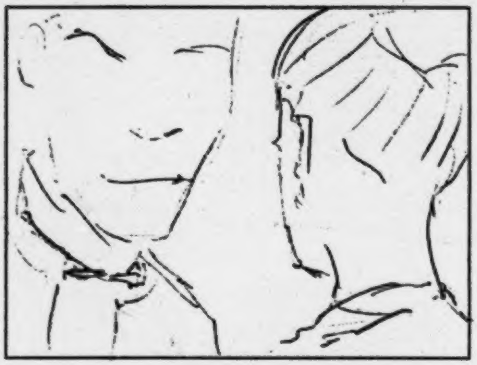
1i



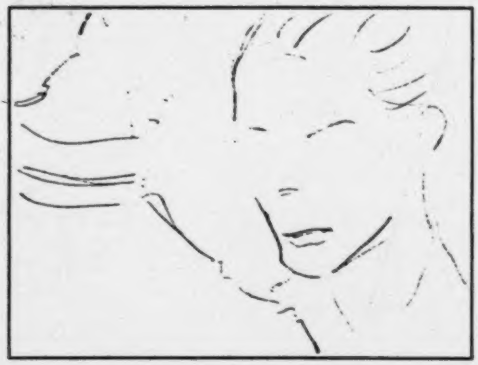
Valmont: "You don't have to speak. Just look at me."

Dangerous Liasions
Seduction #3: 4

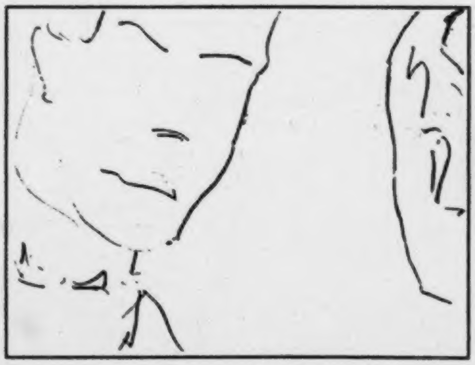
2



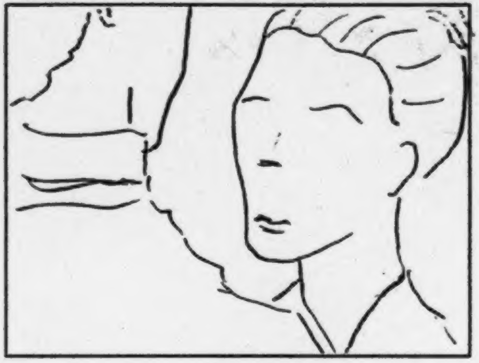
3



4

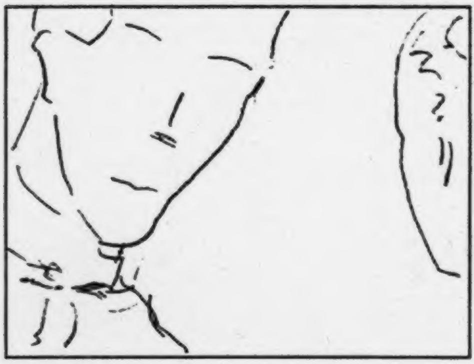


5

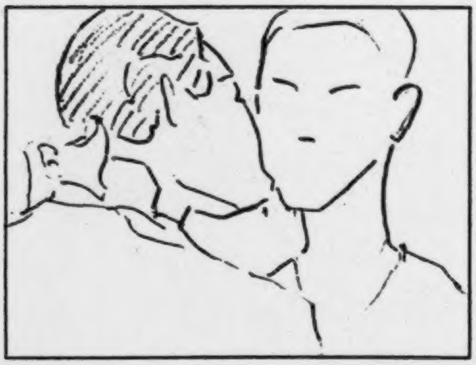


Mme. De Tourvel looks up.
"Yes."

6



7

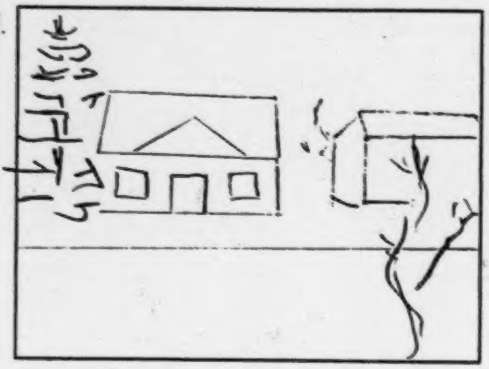


Valmont leans over to kiss her on the mouth, but changes his mind and kisses her on the neck.

Appendix #5

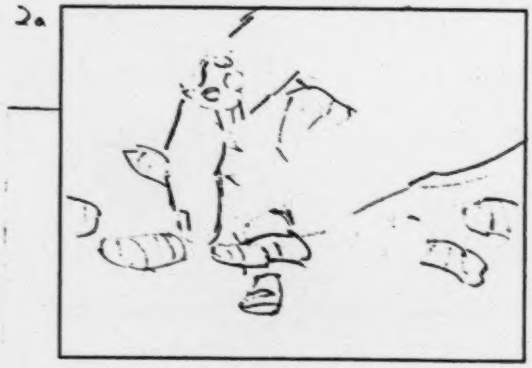
Silences of the Lambs
Surprise at door: 1

1.



Title on image: Calumet City, IL

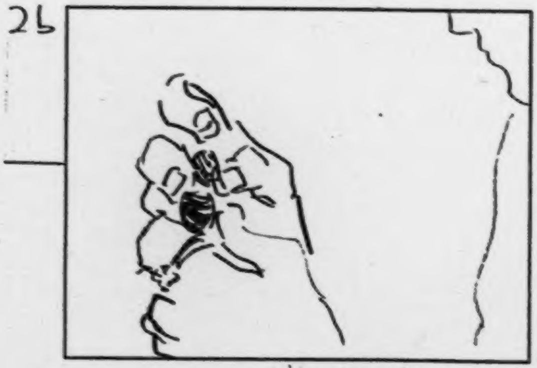
2a



Camera pans right

Hand reaches into bugs' tray.

2b

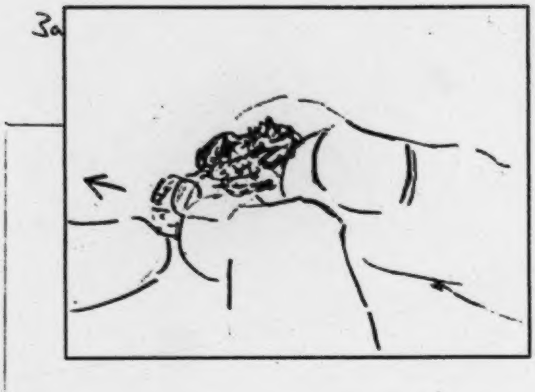


Camera tilts up to [illegible] face of SK

Serial killer (SK) picks-up one bug and looks at it.

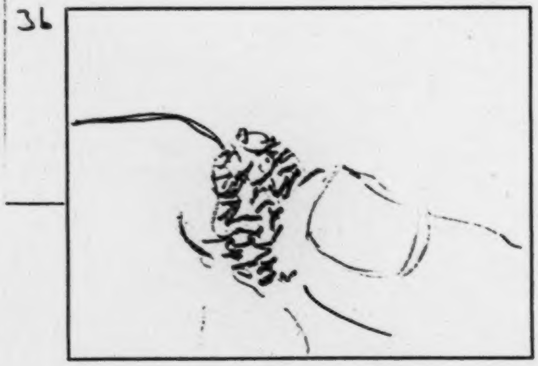
Silences of the Lambs
Surprise at door: 2

3a

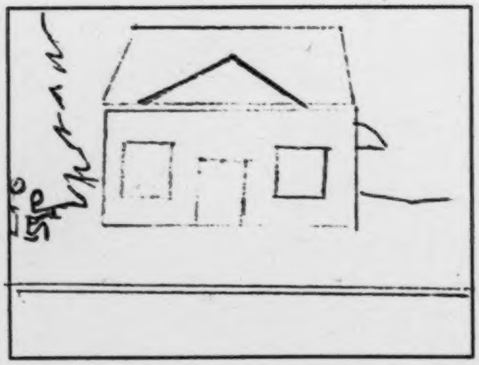


SK pulls out the shell of the bug.

3b

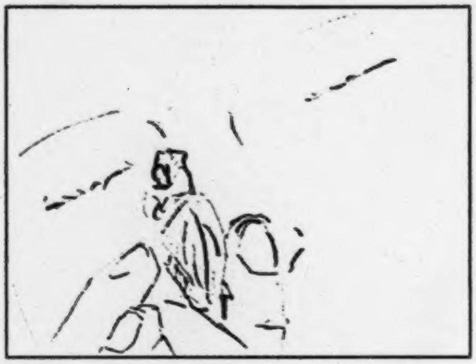


4



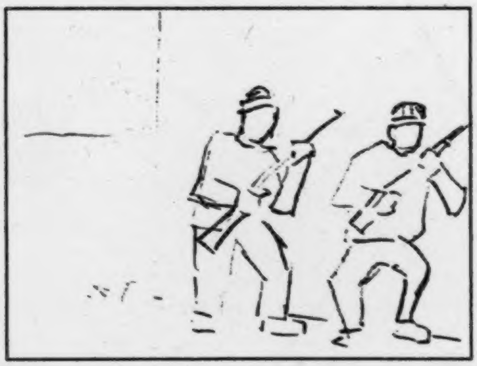
Silences of the Lambs
Surprise at door: 3

5



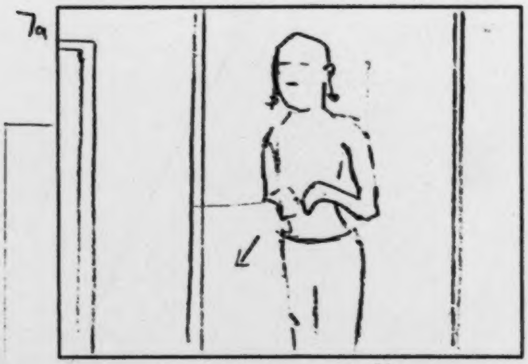
SK: "So powerful. So beautiful."

6



FBI agents approaching.

7a

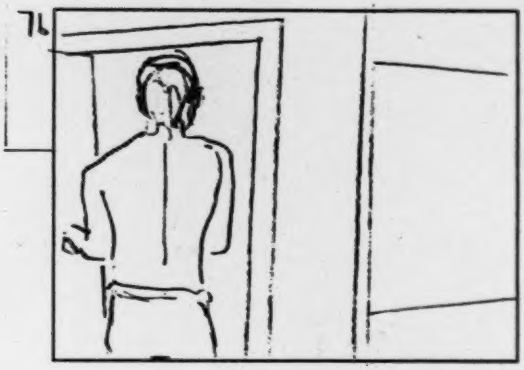


Camera pans left

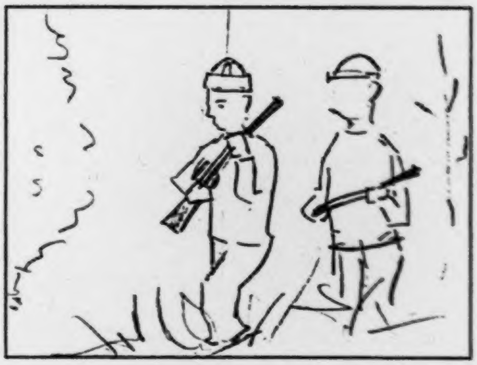
SK (hears something):
"Precious?"

Silences of the Lambs
Surprise at door: 4

7b

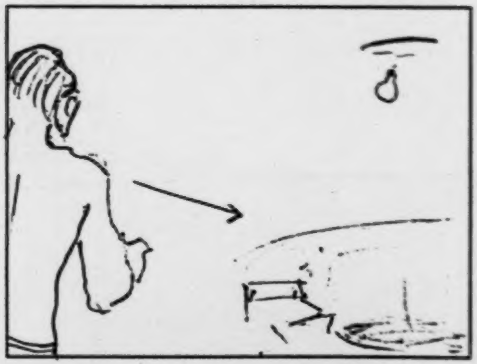


8



FBI agents approaching.

9

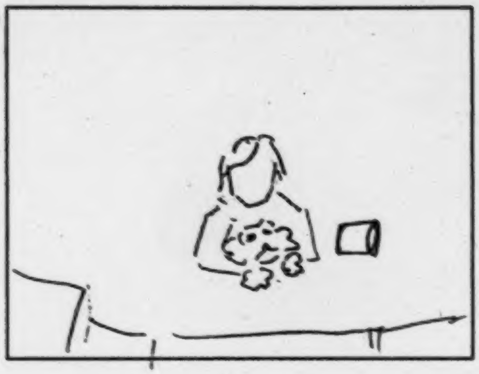


SK: "Precious?"

Camera pans with SK

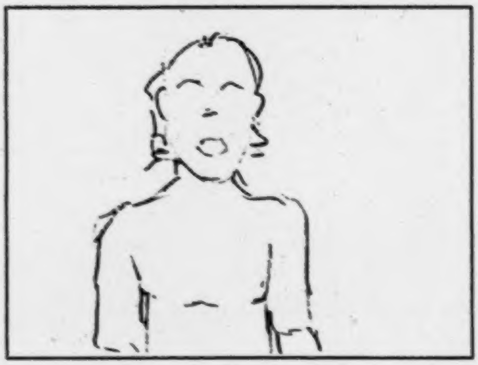
Silences of the Lambs
Surprise at door: 5

10



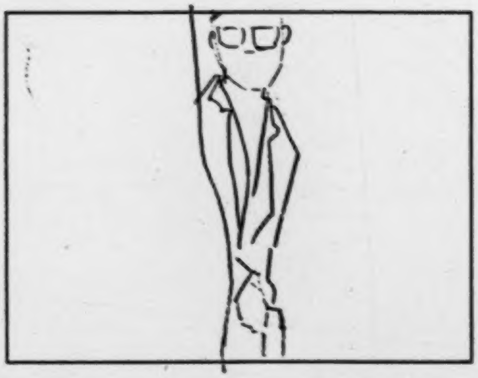
Girl: "She's down here, you sack of shit."

11



SK sighs in surprise..

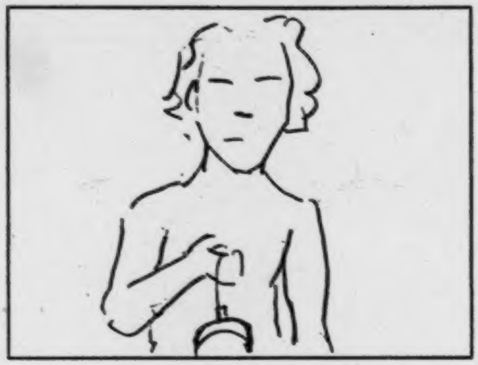
12



FBI superior looks from behind a tree.

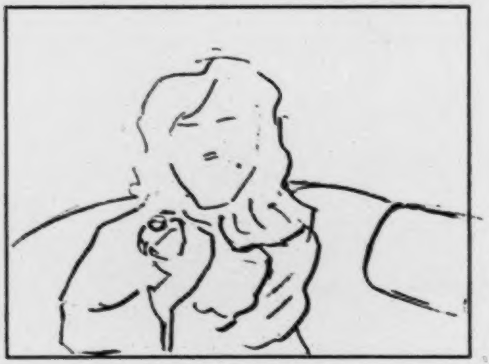
Silences of the Lambs
Surprise at door: 6

13



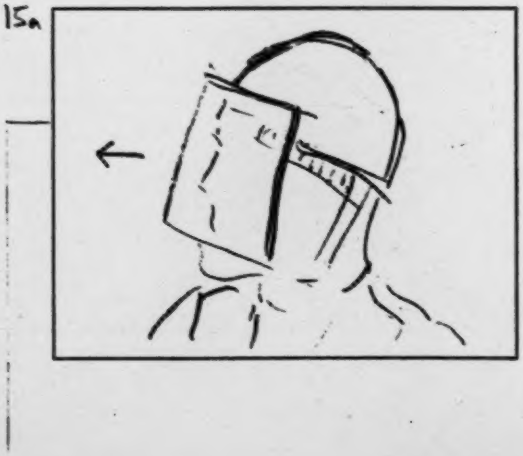
SK (lowers a light): "Put her in that bucket."

14



Girl: "No. You give me a telephone and lower it down here now!!"

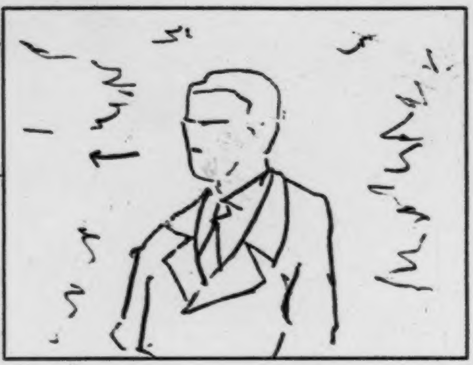
15a



FBI agents moving to left.

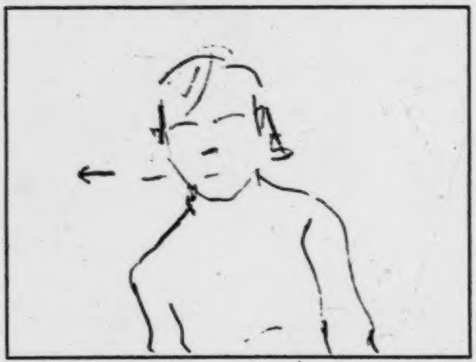
Silences of the Lambs
Surprise at door: 7

15b



Another FBI agents appears.

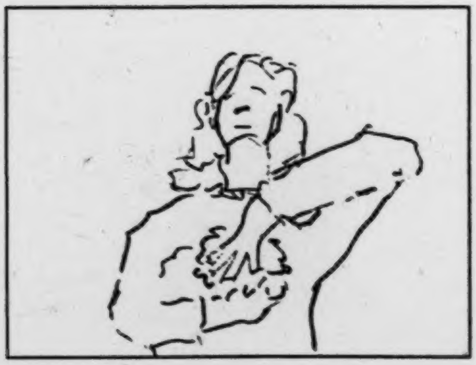
16



Camera pans left

SK (moves left): "Precious, darling, are you alright?"

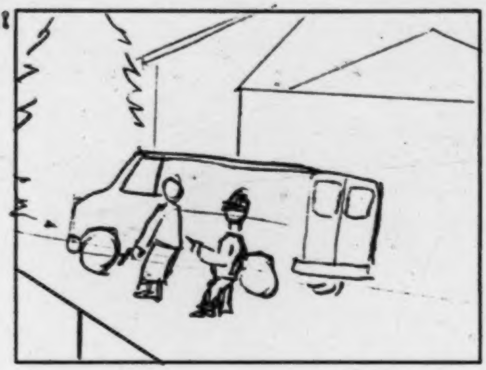
17



Girl: "She is in a lot of pain mister. She needs a vet."

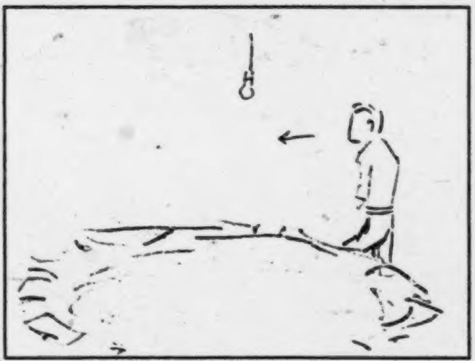
Silences of the Lambs
Surprise at door: 8

18



FBI agents pull a van in front of the house..

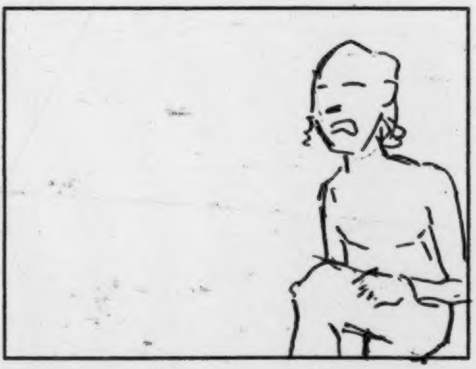
19



Camera pans left

Girl (v/o): She broke her leg. I know it. She's in a lot of pain.

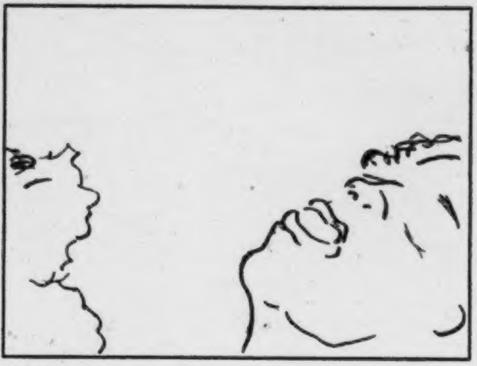
20



SK (yells): "Now don't you hurt my dog!"

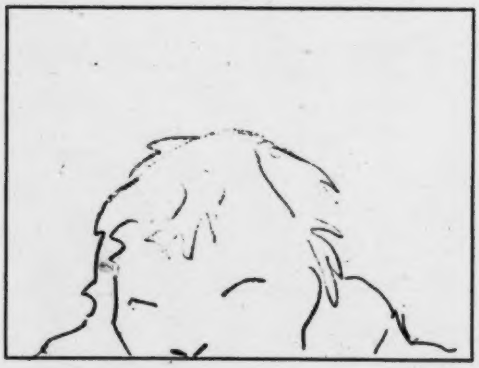
Silence of the Lambs
Surprise at door: 9

21



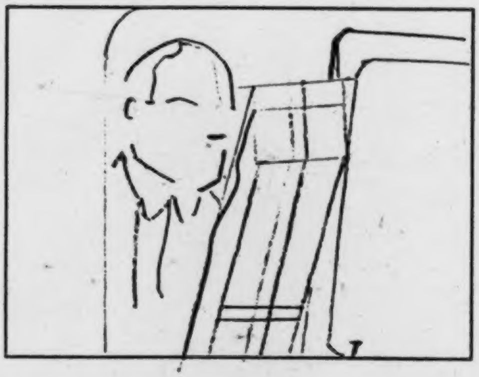
Girl: "Don't you make me hurt your dog."

22



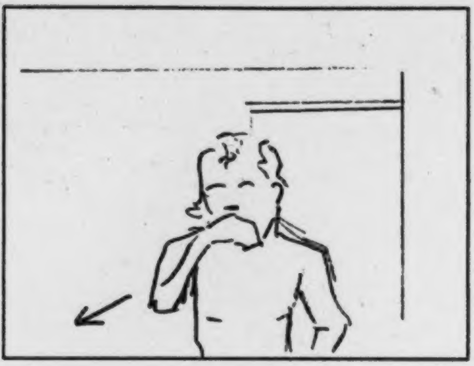
SK: "You don't know what pain is!!!"

23



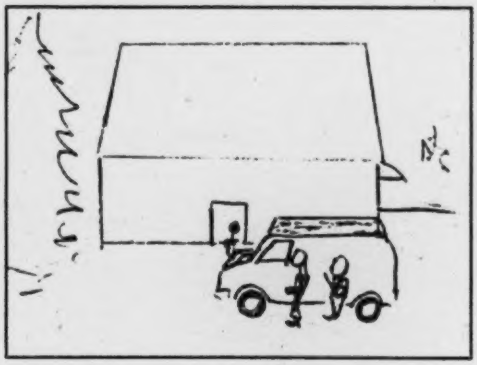
FBI agent pulls a box out of the van.

24

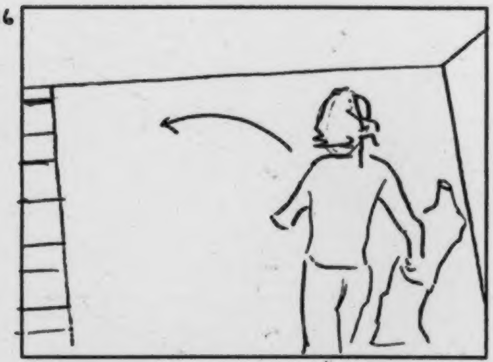


SK is distressed.

25



26

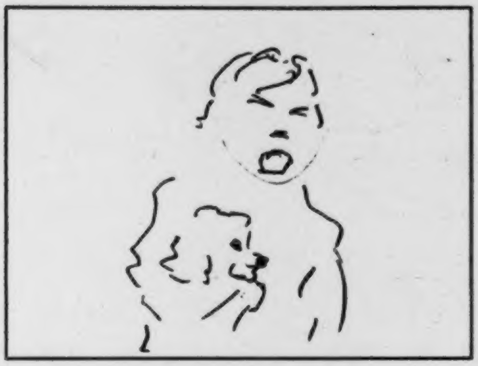


SK trashes room.

Camera pans following SK

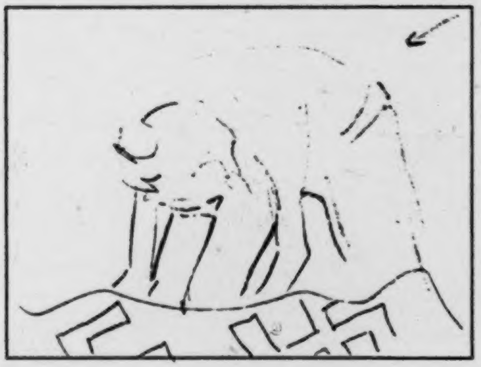
Silence of the Lambs
Surprise at door: 11

27



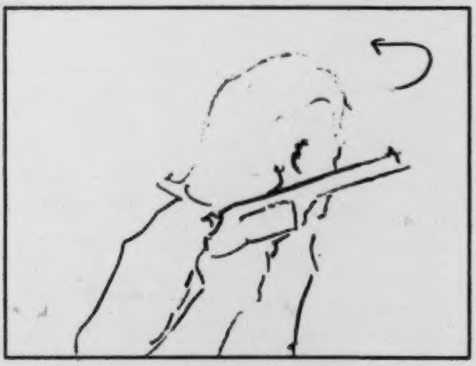
Girl: "I'm gonna do it mister."

28a



SK picks-up a gun.

28b

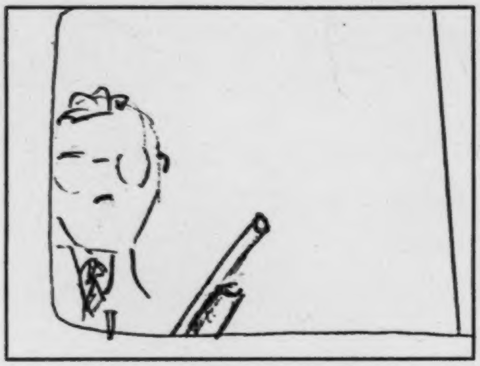


And moves frantically around the room.

Camera pans following SK

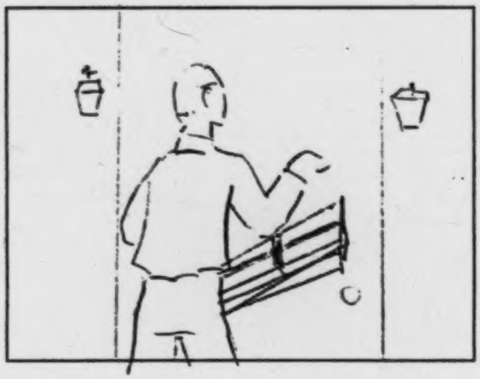
Silence of the Lambs
Surprise at door: 12

29



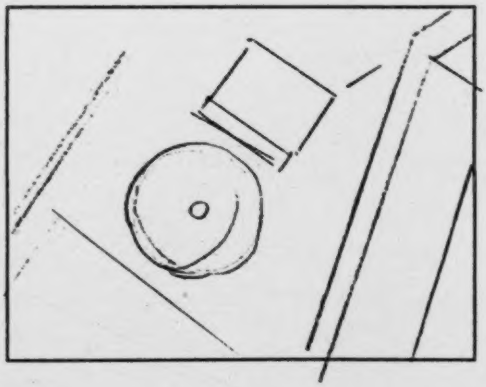
FBI superior gives the cue.

30



FBI agent pushes doorbell.

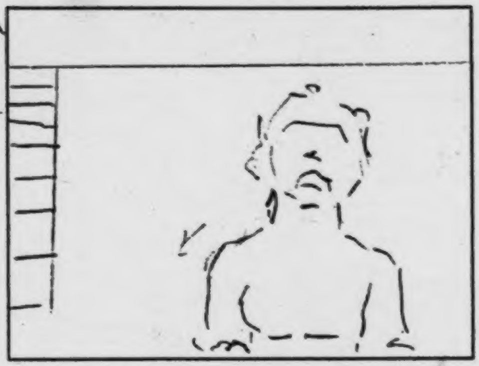
31



Large indoor doorbell rings.

Silence of the Lambs
Surprise at door: 13

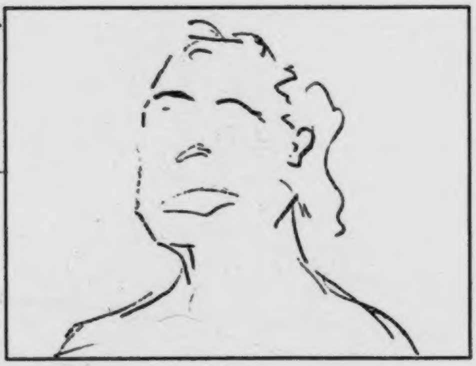
32a



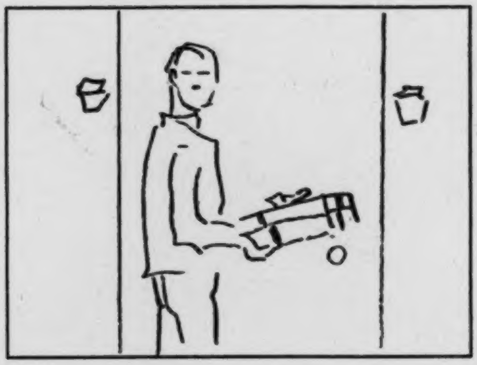
sk moves towards camera

SK listens to doorbell.

32b



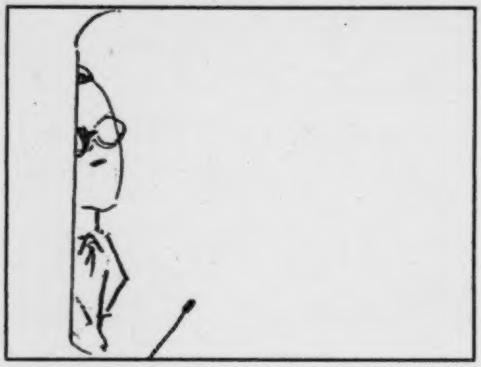
33



FBI agent looks for instructions.

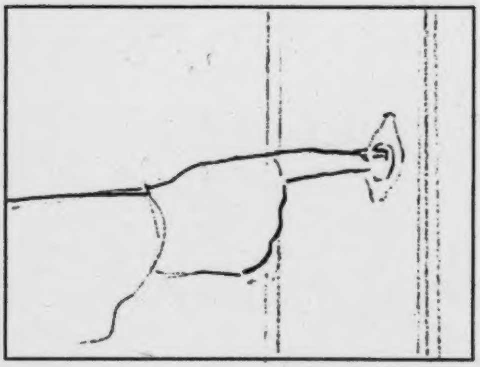
Silence of the Lambs
Surprise at door: 14

34



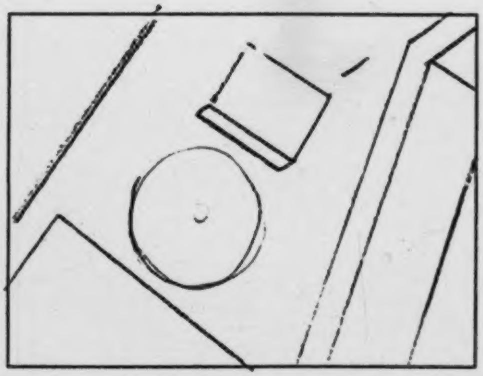
Superior nods yes.

35



FBI agent rings doorbell again.

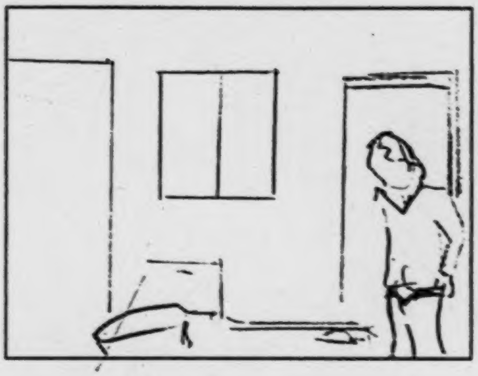
36



Indoor bell rings.

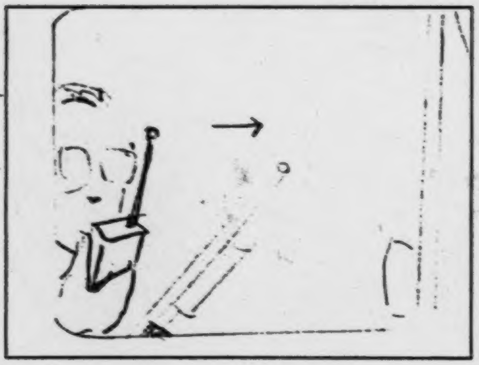
Silence of the Lambs
Surprise at door: 15

37



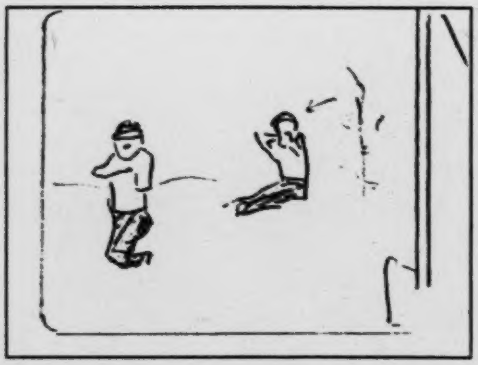
SK opens an interior door, as he is dressing, and mutters: "I'm coming, I'm coming. . ."

38a



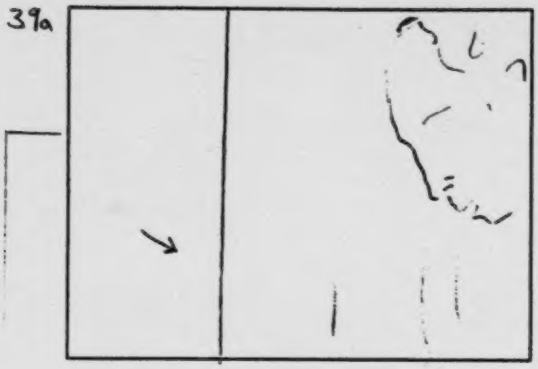
FBI superior: "We're going in." He leaves frame to the right.

38b



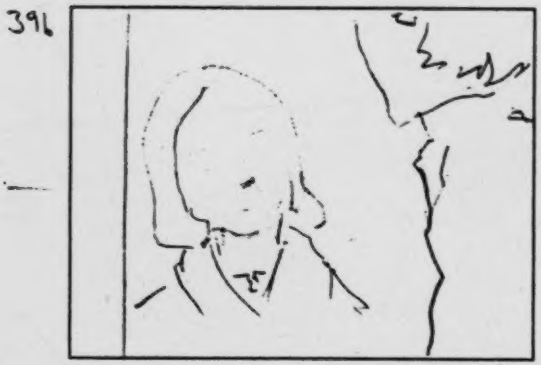
Silence of the Lambs
Surprise at door: 16

39a



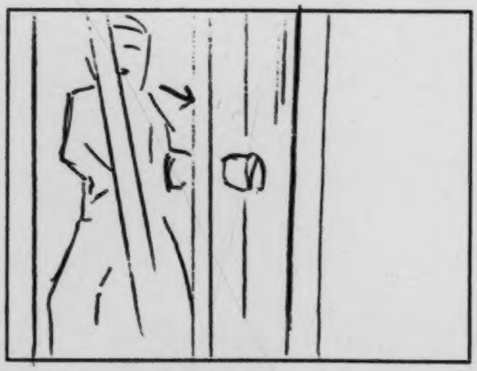
SK opens door to reveal Claris.

39b



Claris: "Good afternoon sir. Sorry to bother you. I'm looking for Mrs. Litman's family."

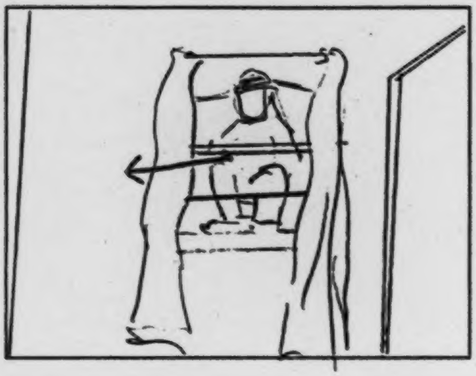
40



FBI agents break in.

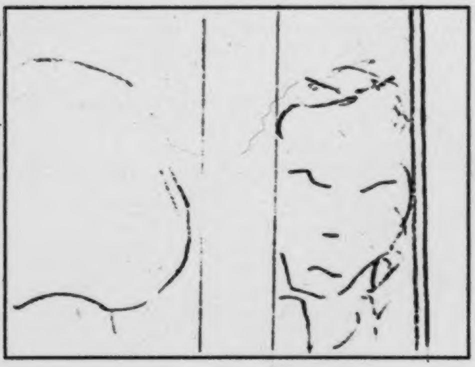
Silence of the Lambs
Surprise at door: 17

41



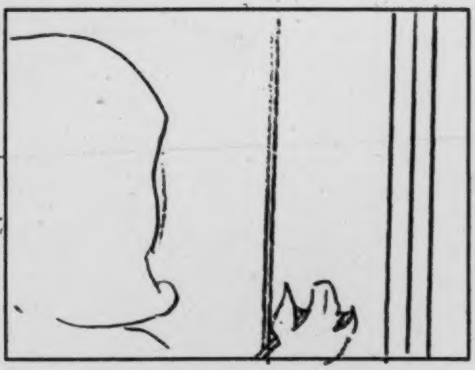
FBI agents break in through the window..

42a



SK: "No Littman's don't live here anymore."
He attempts to close the door.

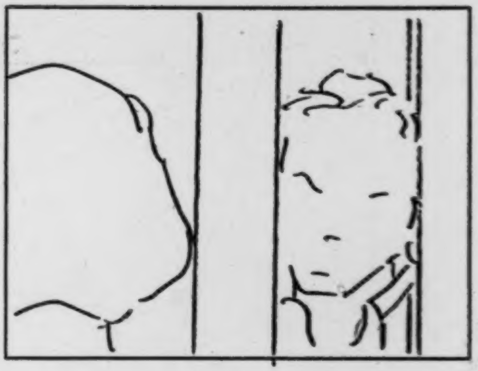
42b



Claris puts hand in door.

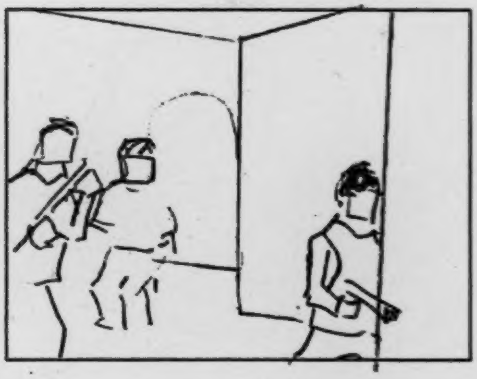
Silence of the Lambs
Surprise at door: 18

42c



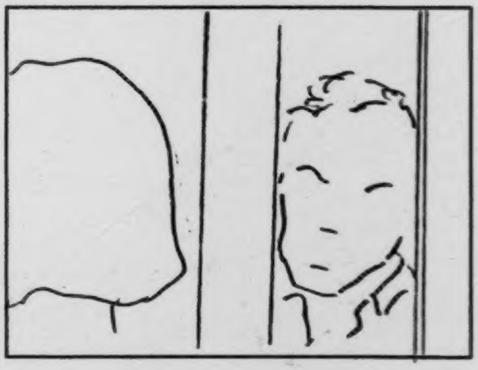
Claris: "I really need to speak with you."

43



FBI agents enter empty house.

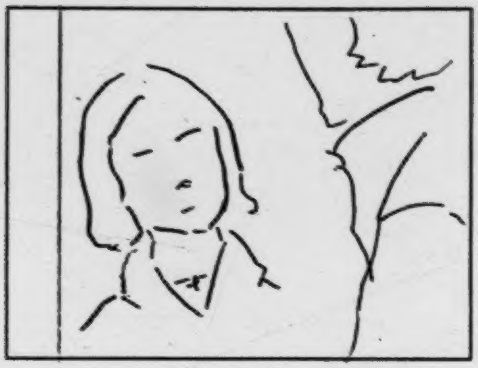
44



SK: "What's the problem officer?"

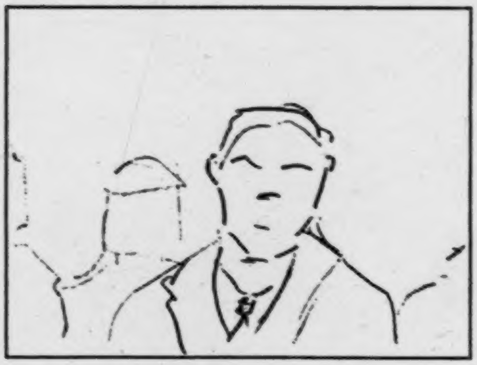
Silence of the Lambs
Surprise at door: 19

45



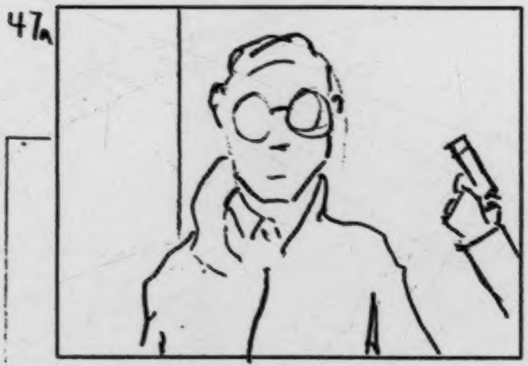
Claris: "Well, I'm investigating the death of Frederica Bimel."

46



FBI agents stand around. The assistant: "There's no-one here, Jack."

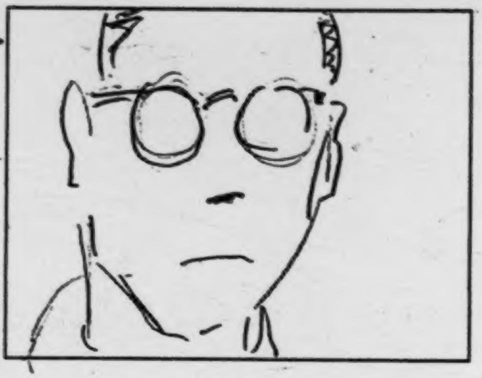
47a



Camera zooms in

Silence of the Lambs
Surprise at door: 20

47b



Superior: "Claris. . ."

48